THE BOOK WAS DRENCHED

RUDIMENTARY MATHEMATICS for Economists and Statisticians

The quality of the materials used in the manufacture of this book is governed by continued postwar shortages.

RUDIMENTARY MATHEMATICS

for Economists and Statisticians

BY

W. L. CRUM

Professor of Economics, Harvard University

First Edition

Second Impression

New York *London*

McGRAW-HILL BOOK COMPANY, INC.

1946

RUDIMENTARY MATHEMATICS FOR ECONOMISTS AND STATISTICIANS

PRINTED IN THE UNITED STATES OF AMERICA

COMPOSITION BY THE MAPLE PRESS COMPANY, YORK, PA.
PRINTED AND BOUND BY COMAC PRESS, INC., BROOKLYN, N. Y.

PREFACE

When the editors of *The Quarterly Journal of Economics* first issued "Rudimentary Mathematics for Economists and Statisticians" as a Supplement to the *Journal* in March, 1938, they did not foresee that a fairly large printing would rapidly be exhausted. Frequent inquiries about the Supplement, since it went out of print, led to the decision to publish the text as a separate volume. In taking this decision, careful thought was given to the choice between retaining the original text practically unchanged except for minor corrections and carrying out a considerable revision in certain parts of the treatment. The second course was chosen, but with the definite understanding that both the plan and the main treatment of the original edition would be retained and that the new materials to be added would be limited with a view to preserving the objective of presenting rudimentary ideas in a treatment directed to beginners. That objective, and the point of view of the treatment, need not now be stated otherwise than as they appeared in the preface of the 1938 Supplement, which is reprinted on page vii.

The present volume could not have been completed so soon, nor could it have reflected such marked improvement in content and presentation, had the task of revision not been largely assumed by my colleague, Professor Schumpeter. His far greater experience with the use and the teaching of mathematics in economics brings this new revision to a higher level than I alone could have achieved. Although we consulted together about all aspects of the revision, the actual preparation of the new materials, both the minor alterations and the substantial additions to topical content, was mainly the work of his hand. So great has been his contribution that I have induced him to permit his name to be included equally with mine in the authorship of the present book.

We are deeply grateful for the interest shown by students and other readers of the *Journal* Supplement and express our thanks especially to those who reported errors that they found in spite of all efforts to avoid slips in the text. We have likewise sought to

avoid slips in this edition. We shall be grateful to readers who call our attention to any that have escaped us or who make other suggestions for the improvement of the book.

So far as the present text reproduces materials that appeared in the Supplement, permission to reproduce has been granted by *The Quarterly Journal of Economics;* and the authors are grateful for this generous permission.

W. L. CRUM.

CAMBRIDGE, MASS.,
March, 1946.

PREFACE TO THE FIRST EDITION

The objective of this book is to present rudimentary ideas and operations essential to any effective mathematical reasoning by economists and statisticians. The book does not aim to present a systematic treatise on what has come to be called mathematical economics, nor is it a systematic treatise on elementary mathematics or even a comprehensive discussion of the numerous mathematical analyses of elementary economic theory or statistics. It is directed to the mere beginner—to the reader who has never studied mathematics beyond the first-course stage of algebra and geometry or has studied "advanced" mathematics so long ago that his memory is hazy or blank.

The present-day economist feels an impelling need for mathematical understanding. Published economic work, in books or learned journals, places an increasing reliance upon mathematical methods of analysis; and the reader who does not comprehend such methods is under a severe handicap. Likewise the classroom student of economics—graduate and undergraduate—is constantly encountering subject matter in which mathematical methods are helpful or essential. The necessity for removing, at least partially, the mathematical deficiency under which many economists labor is the occasion for this book. The book is not a substitute for systematic training in mathematics; and, in offering the book to those economists who may hope to benefit by its study, I shall not relax advocacy of every effort directed to extending and strengthening the mathematical training of prospective economists. I surely express the sincere wish of many colleagues in venturing to hope a later generation of economists will be so well prepared in mathematics that occasion for a book of this sort may disappear.

Long observation of the mathematical difficulties of my students has given me a deep-rooted conviction that their mystification results almost invariably from a lack of precise understanding of the very rudiments of the subject. The "mystery of the calculus" is a mystery largely because students have never really

mastered the central ideas from which the calculus is developed or ideas still farther back and still more basic than those central ideas, or because such understanding as they have had runs in terms of physical and geometrical illustrations that cannot readily be identified with the problems of economics. At the risk of giving a presentation that may seem painfully detailed and monotonous to the reader, I have written these pages with the precise purpose of schooling him so thoroughly in certain root ideas that he can proceed with self-reliance to the more complex problems and elaborate treatments found in various texts and other published work. To the student who will faithfully and patiently master these pages, subsequent mathematical work in economics will appear less clouded in mystery. That such subsequent work may involve many concepts and operations not treated in the following pages is gladly granted; the book aims merely to provide that solid foundation of elementary knowledge which will give the student self-reliance to erect a superstructure of more advanced knowledge by his own efforts.

The material of the book has been organized to develop the mathematical topics from the simplest types of charting of elementary algebraic expressions, and related notions of analytic geometry, through the central and vital treatment of limits and rates and derivatives, up to a brief consideration of the location of maxima and minima and the formulation and solution of simple problems by the method of differential equations. As mathematics is usually taught, the passage from geometry and algebra to differential equations is by a fairly long route; and the compression attempted in this book has been undertaken deliberately, and with a conviction that it is feasible.

Throughout, the illustrative material has been drawn almost exclusively from economic theory, with a few cases from statistics, in the belief that one of the chief difficulties of economists is the expression of their problems in symbolic terms. Naturally, the illustrations cannot touch in any comprehensive way the great range of mathematical applications in economic theory; and the specific illustrations are necessarily at times simplified by omitting factors or qualifications essential to careful economic reasoning. At frequent points where such oversimplification has seemed wise, I have noticed the fact as a warning to the reader. Likewise, in the mathematical argument, I have at various points indicated that

important questions of refinement in reasoning have been ignored. While the mathematical analysis is intended to be valid as far as it goes, careful and complete insistence on rigor has been neglected lest the beginner suffer an even greater bewilderment than that which the book seeks to remove.

Great care has been taken to ensure accuracy, in both the mathematics and the economics; but I doubt not that various imperfections remain. I shall be grateful to readers who call my attention to errors or make other helpful suggestions in anticipation of the uncertain day when another edition may be attempted.

W. L. Crum.

Cambridge, Mass.,
March, 1938.

CONTENTS

RUDIMENTARY MATHEMATICS FOR ECONOMISTS AND STATISTICIANS

CHAPTER I

GRAPHIC ANALYSIS: SIMPLEST CASE

Numerous problems in economics and statistics require the study of items—such as output, cost, wages, rent, profit, investment, income—that vary, either as time passes or from place to place or case to case; frequently the variation of such an item has a form that implies the existence of some limiting size for the item. The determination of this limiting size, and the way in which the limit is approached, is the key to much of the content of this book. For the study of limits, relatively simple cases from the theory of costs provide useful illustrations. After a rather detailed examination, in this and the next chapter, of the graphic method and of very elementary matters of definition and principle, we shall proceed in the third chapter to develop the limit concept.

Total Cost. Suppose that a single producer is turning out a single and uniform commodity, under conditions of cost that are to be examined. When his operations are at some particular rate, let the quantity of output per unit of time be q and the corresponding total cost of producing this specified output be c.[1] The first task is to develop a mathematical representation of the relation between q and c for each of several simple assumptions concerning the nature of his costs.

The desired mathematical representation takes two forms, the graphic and the symbolic. The graphic representation is probably already familiar to the reader, from elementary statistics. Imagine that the producer has compiled a table of the

[1] Thus q—and the same is true of c—is a *rate* of the sort called a "time rate." The rate idea is of basic importance in economic analysis, as will appear below (Chap. IV), but it need cause no worry at this stage.

cost c associated with each of several quantities of output q. He may have secured this table by calculating c by adding, for each selected q, the various elements (constituents) of cost—for labor, materials, use of equipment, supervision, and so on—that he or his engineer or other expert knows or estimates would be involved in producing q. This method is a priori or *deductive:* it arrives at c by starting from certain known or assumed rules or laws that connect the constituents of cost with the volume of production.

On the other hand, he may have secured this table by observing c for each of the actual quantities of output specified by q. Such an observational or experimental determination of the table should, to be satisfactory for the present discussion, take place under conditions that economists and other scientists, sometimes loosely and perhaps often lazily, describe as "all other things

TABLE 1.—TOTAL COST c INCIDENT TO PRODUCING EACH SPECIFIED QUANTITY OF OUTPUT q OF A UNIFORM COMMODITY BY A SINGLE PRODUCER WITH UNCHANGING CONDITIONS OF PRODUCTION*

q	c	q	c
0	16	5	56
1	20	6	70
2	26	7	86
3	34	8	104
4	44	9	124

* **Units:** For q, thousand yards; for c, hundred dollars.

being equal." What does this vague phrase mean here?[1] It means that, as the producer passes from one stage of his experiment to another, q is the *only* variable magnitude, affecting his costs, that he allows to change. Other variables that might affect costs—such as the price of materials, the type and arrangement of his machines, the wage scale, the skill and discipline of his labor, the supervisory and other organizational aspects of production—remain unchanged.[2] All this does not imply that

[1] We dwell upon this point partly because of the intrinsic importance to the present discussion, and partly to accustom the reader to the necessary habit of questioning the precise significance of every word and phrase.

[2] This presupposes that the producer is operating under conditions of the simplest and most perfect competition; thus prices of the factors of production do not change as he expands his output.

the methods to be studied below can never be adapted to cases in which one or more variables other than q may change. It merely means that, for the present purpose, a simple and effective treatment requires the experiment to be conducted with only the single variable q changing and bringing about changes in c. If the table has been arrived at by this observational procedure, the data in that table are called *empirical* and the process of drawing inferences from the data is *inductive.*

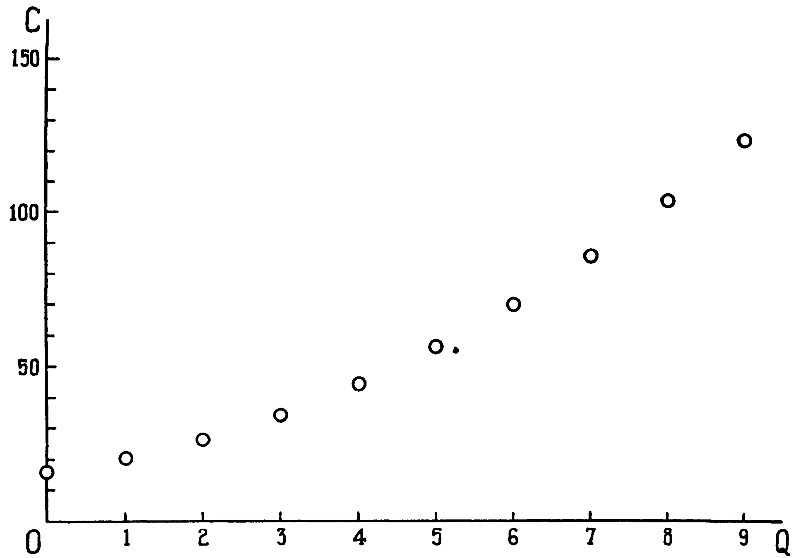

CHART 1.—Costs of particular outputs, based on Table 1.

Whichever method the producer used in obtaining his figures, let them be as given in Table 1. We undertake first the graphic representation of these facts (Chart 1). Measure a particular value of q in the horizontal direction indicated by the axis OQ, from O as zero, in some scale, relating distances measured along OQ to specified sizes of q, to be determined.[1] Likewise, a particular value of c is measured in a vertical direction indicated by the axis OC, from O as zero, in a scale to be determined. Actually, a line (vertical) is drawn parallel to OC at the distance q, as above measured, from OC; c is then measured, upward

[1] The direction is a matter of convenience or custom: q might have been taken vertically and c horizontally.

along this line, from OQ. This process locates the desired point. The horizontal measurement from O is called the *abscissa*, and the vertical measurement the *ordinate*, of the point to be located. The pair, abscissa and ordinate, are called the *coordinates* of the point. Horizontal distances toward the right are taken positive, toward the left are negative; vertical distances upward are taken positive, downward are negative. There is no fixed rule for choosing scales, the number of 1,000-yard units of q per inch along OQ, or number of \$100 units of c per inch along OC; the choice is a matter of convenience, largely governed by the desired size of the diagram. Moreover, there is no necessary connection between the scale for q and that for c (see, however, page 10), again it is a matter of convenience; a common practical rule is to choose the two scales so that the completed diagram will have a height somewhere between 60 and 100 per cent of its width.[1]

Having reached the foregoing decisions as to how to make measurements and as to the scales, a point can actually be located on the chart to represent each of the 10 pairs of numerical magnitudes, for q and c, given in Table 1. The result is shown in Chart 1. Here the points (represented by small circles, to make them clearly visible) are not joined by a curve, or by a series of line segments. The reason for refraining, temporarily, from doing this is that Chart 1 is designed as an exact representation of the record given by Table 1. Had a curve been drawn through the points, or had successive pairs of points been connected by straight lines, the chart would have conveyed an impression as to costs for other values of q than the particular cases specified in the table. For some purposes, as we shall see, this is desirable and justifiable; but, for the moment, the chart gives no information except that in the table.[2]

[1] In practice, one ascertains the range between the lowest and highest values of q, and likewise for c, and bases the scale selection on these. Or, if one wishes to show the zero point of the chart, as in the present case, one uses merely each highest value to guide the selection of scales. This independent choice of scales for q and c is satisfactory for the mere graphic presentation of the data, as in Chart 1; for certain analytical purposes, the scales cannot be independent. For example, see footnote 1, p. 25.

[2] Anyone accustomed to study charts will almost surely form in his mind an imaginary picture of just such a curve as we have purposely omitted and proceed to draw inferences from it concerning the relation between q and c

The graphic representation in Chart 1 conveys to the mind of the reader certain impressions concerning the relation between q and c; some of these are conveyed also by Table 1, but some of them are afforded not at all by the table, or only imperfectly and to a reader well skilled in understanding tabular data. Manifestly there is some cost, c being 16 for q zero, even when

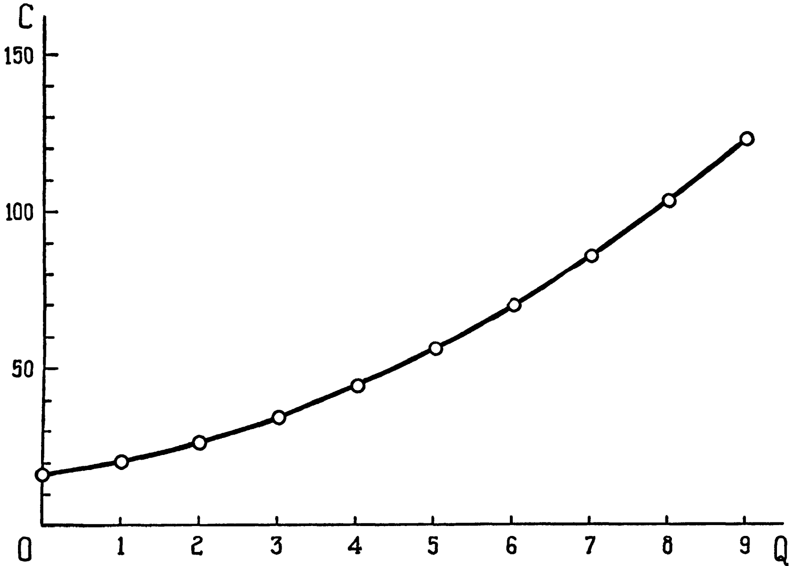

CHART 2.—Continuous cost function inferred from data of Table 1.

operations are completely stopped. We observe also that the points slope upward as our eyes pass to the right: total cost increases as output increases, which is what we should expect. We note further that the points rise more rapidly as we pass to the right, that the vertical movement from one point to the next becomes greater as we pass to the right: cost not only increases with output, but increases more rapidly than output.[1]

---

at points intermediate between those of the table, or perhaps to the right of the final point where q is 9. Correspondingly, we sometimes say in elementary statistics that such a curve can properly be drawn "merely to guide the eye from point to point"; but this intent to restrict the use of the line will not prevent the reader of the chart from using it also as an aid to *interpolation* (the estimating of intermediate points).

[1] This fact could also be discovered from the table by carrying out a small operation on the data: the *first differences* of c, obtained by deducting each c from that next following it in the table, increase as q increases.

This third observation is more clearly indicated if the points are joined by a curve. This is done, waiving the objections raised above as to danger of interpolation, in Chart 2. There the curvature—the nonlinearity—of the array of points is emphatically apparent. Moreover, the curve is concave upward: its hollow side is above the curve. This fact shows definitely that total cost increases more rapidly than output, but we must wait until later (page 119) to reveal this comparison in quantitative terms.

Cost Proportional to Output. Before further analysis of the given record of costs for the single producer, certain additional concepts can helpfully be introduced. To this end, a succession of particular assumptions concerning costs is needed; these assumptions are of the a priori sort—it is assumed that the law relating cost to quantity is known, without experimentation.

The first assumption is that a producer operates under conditions such that total cost is directly proportional to output. Although in actual economic life such a case of costs is not easily found, because even the simplest type of cost, mere unskilled labor, is likely to involve questions as to whether mere addition of more labor effectively increases output in the same proportion, the student can readily imagine cases in which these conditions might be nearly realized. Even though it is not consistent with economic reality, the problem in the form stated has certain advantages for purposes of analysis preliminary to the study of more realistic situations. On the assumed basis, if output is doubled, total cost is doubled; if output is tripled, total cost is tripled; and so on. Elementary algebra teaches that this very simple law, connecting quantity and cost, can be expressed by the equation

$$c = kq \tag{1}$$

where k is a constant—does not change when q changes. In fact this equation may be regarded as a definition of the concept of direct proportionality. For a particular value of q, the formula, assuming k is known numerically, gives the corresponding value of c.[1]

[1] The term "value," as used in mathematics, refers to the size (numerical size, with due regard to + or − sign) of a magnitude represented by a symbol. The student of economics, of course, uses the word value in a different

In Equation (1), c and q are *variables*, the former generally called the *dependent variable* and the latter the *independent variable*, and k is a *constant*, which means that it is a fixed or unchanging magnitude regardless of the various values taken on by the variables q and c.[1]

Equation (1) is said to express a *functional relation* between q and c; or, alternatively and more often, c is said to be a *function* of q. For the present, a sufficient definition of this concept is: If the value of the variable c depends upon, or is determined by, that of the variable q, c is said to be a function of q. Furthermore, the functions that ordinarily concern us in economics are single-valued: for each value of q there is only one value of c. In Equation (1), c stands completely alone—to the first power, and without any multiplier or any other mathematical symbol attached to it—on one side (usually, as here, the left) of the equation, and the other side (usually, as here, the right) of the equation is an expression not involving c but involving only q and the constant. Such an equation is said to give c as an *explicit* function of q. The expression on the right side might have been much more complicated than that of Equation (1); it might, for example, have included the sum of several terms, containing various powers or roots or other functions of q, with or without multiplication by various constants. But, so long as c did not appear on the right side, and appeared completely alone, as above, on the left side, the equation would give c as an explicit function of q. Thus

$$c = aq^3 + kq^2 - bq$$

gives c explicitly in terms of q and the constants a, k, and b (the constants are assumed chosen appropriately).

---

technical sense. He will have no difficulty in identifying which of these two technical significations is implied in the various instances in which the term is used in the present text. The point is noted here merely to put him on his guard against confusing the technical meanings of any term having more than one such meaning.

[1] Which variable, in a problem involving only two variables, shall be called dependent and which independent is, from a mathematical point of view, a matter of convenience. In the present instance, we call c dependent because we are thinking of trying different outputs q and then discovering the corresponding costs c. But, so far as the symbolism is concerned, it makes no difference.

In the terminology of elementary algebra, such an equation has been solved for c in terms of q. On the other hand, an equation like

$$ac - c^2 = kq^2 + cq + 3q$$

is said to express c as an *implicit* function of q (although in this case c is not a single-valued function of q): c depends upon q, but the equation has not been solved for c in terms of q.[1]

Equation (1) is an example of symbolic representation and has certain notable advantages over the graphic representation utilized in the preliminary study of Table 1. One of the advantages of such a representation is its generality, a property contributed by the unknown, but presumably determinable, constant k. In fact, Equation (1) is equally valid for every producer for whom the proportional law holds. For one such producer cost may be three times output, for another five times output; but these differences can be cared for by assigning different numerical values, 3 and 5, to k. Thus, the only difference among producers is that different numerical values of k apply, or may apply, to them.

In this sense, that is to say, if we are considering different producers, k is also a variable. Such magnitudes, which we treat as constants when dealing with one aspect of a problem and as variables when dealing with another aspect, we call *parameters*. Another example may be useful. The student is likely to come across the proposition that, under conditions of perfect competition, prices "play a parametric role." This means that in a perfectly competitive market no buyer and no seller is in a position to influence prices by his own single-handed action (see footnote 2, page 2). Prices are therefore data for every individual buyer or seller, and we accordingly treat them as constants when describing individual behavior in a market; whereas for any other purpose they are of course not data but, on the contrary, figure among the most important variables that economists seek to determine.

[1] In all elementary work, we strive to avoid implicit functions; often, by solving for one variable in terms of the other, we can convert an implicit function into an explicit function. Some implicit functional relations cannot thus be solved, but the student should not infer that this fact necessarily raises an insuperable obstacle to analysis.

Although a considerable analysis of the general equation

$$c = kq \tag{1}$$

would be possible, and would yield much helpful information, present purposes are better suited by first adapting the formula to a single producer—by rendering the general formula specific. The customary device for rendering a general formula specific is to impose an initial condition.[1] This consists in substituting a known pair of values for q and c in the formula and determining k so that the equation will be satisfied. If we know, for the single producer, the following pair of values

$$c = 6, \qquad \text{when} \qquad q = 2$$

k is 3, as found by substituting the known values of q and c in (1). The specific formula for the single producer is thus

$$c = 3q \tag{2}$$

which enables us to calculate c for any value of q.

Average Cost. The average cost per unit for a particular output q is defined as the ratio of c to q, and we designate it by the symbol $\bar{c}$. In most cost problems the words "for a particular output" are essential to the definition, as the average cost varies for various outputs; but in the highly special case under examination in the preceding section, as will appear immediately below, the average cost is constant for all values of q and the qualifying phrase is needless.[2] Thus, in general

$$\bar{c} = \frac{c}{q} \tag{3}$$

[1] The word "initial" is unfortunate, because the condition imposed need not, as appears in the text, have anything to do with a "beginning." It can apply to any known pair of values of q and c (see p. 141).

[2] The words "per unit" are in a sense redundant; they are implied in the word "average" as here used and as ordinarily used in elementary economic theory. We shall learn, as we proceed, that many technical concepts are not absolute but relative, and require qualifying phrases in order that their significance shall be precisely known. In the present instance, as remarked, the words "per unit" are not essential, though they may protect us from the error of thinking that we refer to the average "over a period of years" or "among several producing firms."

For the case of total cost directly proportional to output, discussed in the preceding section,

$$c = kq \tag{1}$$

and, hence, by (3)

$$\bar{c} = \frac{kq}{q} = k \tag{4}$$

For that case, then, the average cost is identical with the constant k; and, with our present definition of average cost, we might have written the general Equation (1) in the form

$$c = \bar{c}q$$

specifying that $\bar{c}$ is fixed or constant.

Likewise, for the single producer for whom k is 3

$$c = 3q \tag{2}$$

and

$$\bar{c} = 3 \tag{5}$$

Certain advantages will arise from turning aside now to study charts of Equations (2) and (5). Students familiar with analytic geometry will know that Equation (2) represents a straight line. Elementary geometry teaches that "two points determine a straight line." We therefore find two pairs of values of q and c that satisfy Equation (2). To do this we select two separate numerical values of q, substitute them in Equation (2), and find the corresponding values of c. For example, one such pair is (0, 0) and another is (4, 12).[1] By plotting these two points on a diagram, with axes of measurement chosen exactly as for Chart 1 but scales in this case determined so that a unit of q is measured (for reasons given below, page 14) by the same distance as a unit of c, and then ruling a straight line through these two points, we have a complete graphic representation of Equation (2), see Chart 3.[2] In ruling the line, although strictly an endless straight

[1] In stating a pair of values for two variables, it is customary to list them in parentheses, the independent variable being listed first. Likewise, the coordinates of a point are listed in parentheses, the abscissa first and the ordinate second.

[2] The student not familiar with analytic geometry will want evidence that the straight line shown does in fact represent Equation (2); to date, all he knows is that two points of the line (0, 0) and (4, 12) belong to Equation (2). In the jargon of the mathematician, such evidence, to be conclusive, consists in showing that the line is the *locus* pertaining to the equation, which means that the coordinates (1) of every point on the line satisfy the equation and

line passes downward to the left of O as well as upward to the right, we do not pass left of and below O; economic considerations obviously exclude negative output and negative cost as having no significance.

The line in Chart 3 truly and completely represents, in graphic form, the relation between cost and output. Unlike Chart 1, in which the plotted points represented only isolated pairs of values of q and c as given in Table 1, this chart shows a continuous line that represents the relation for every value of q, not only whole numbers, but fractions. The appropriateness of such a continuous representation might be more apparent if we were considering such a highly divisible commodity as a liquid in bulk.

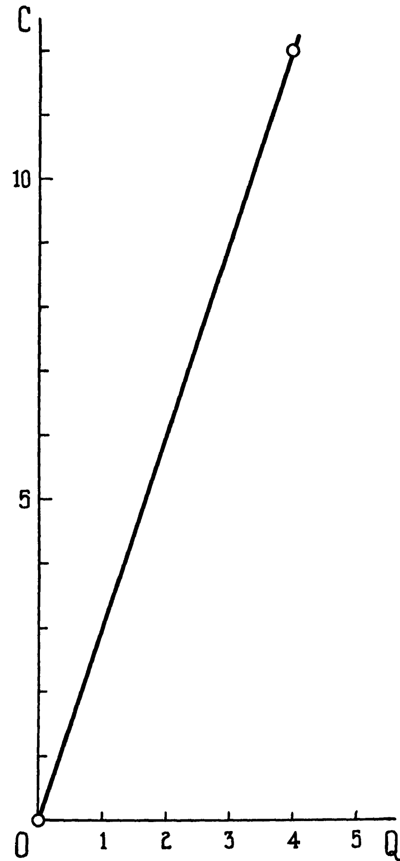

CHART 3.—Continuous cost function for total cost proportional to output.

As a means of describing the line, we can of course report the positions of any two points, the two plotted, or any others, and rely upon the fact that two points determine a line. For some purposes, a different description of the position of the line is more helpful. This description customarily consists in reporting one point on the line—in this case the most convenient such

(2) of no point not on the line satisfy the equation. We shall not include a complete proof of facts 1 and 2 for this line, but merely suggest that the student can largely remove any doubts he entertains by making tests: Choose any point on the line, measure its q and c from the chart, and verify that they satisfy (at least as closely as the accuracy of his measurements permits) the equation; choose any point off the line, and verify that its q and c do not satisfy the equation.

point is (0, 0), but any point would do as well—and the direction of the line. The direction may be specified in either of two ways, by the inclination or the slope; to facilitate definition of these terms, the line is reproduced in Chart 4. The *inclination* of the line OA is the angle that it makes with the horizontal axis OQ, such angle, designated by the symbol α, being measured positively from the positive horizontal axis OQ counterclockwise to the line OA. As a special case, the inclination of a horizontal line is zero. The student will know from his elementary geometry that this same angle α would be found if the measurement were made from the positive direction of any horizontal line—prolonged, if necessary, to intersect OA—and not exclusively from OQ. Just as the two points (0, 0) and (4, 12) completely determine the line, the one point (0, 0) and the angle α completely determine it; no other line meets these requirements.[1]

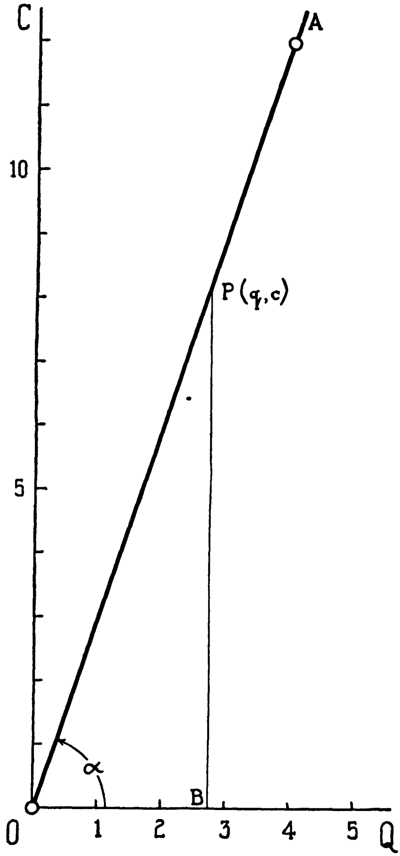

CHART 4.—Angle of inclination.

The *slope* of the line OA involves a concept from trigonometry, the tangent of the angle α.[2] The slope of a line is defined as the

[1] For the particular line as shown, α of course has a definite numerical size. We might measure it with a protractor, finding it about $71\frac{1}{2}°$. A more precise numerical determination will be given immediately below.

[2] The *tangent* of an angle—the student must not confuse this use of the word tangent with that implied in "tangent to a circle" or to some other curve—is usually defined in terms of an angle smaller than 90°, an acute angle. If such an angle is regarded as one of the two acute angles of a right triangle, its tangent is the ratio of the length of the side of the triangle oppo-

tangent of the line's inclination. The tangent of α is BP divided by OB. P is a general point—any point, other than O—on line OA; the student will note that these label letters have carefully been read away from the respective axes and toward the point. For the present case, these are positive directions; but if P were located so that q and c were not both positive, reading in this way would yield the correct sign for tan α. As P is (q,c), the tangent of α, customarily represented by the compound symbol tan α,[1] is

$$\tan \alpha = \frac{BP}{OB} = \frac{c}{q}$$

and, as q and c satisfy Equation (2)

$$\tan \alpha = \frac{3q}{q} = 3$$

The point (0, 0) and the slope 3 completely determine the position of the line. Knowing that tan α is 3 enables us to determine α in angular measure more precisely than by a protractor. Thus, a table giving tan α for various values of α shows that, when tan α is 3, α is $71°45'$ approximately.

We might infer that for the more general case of Equation (1), which we may suppose graphically represented by a line of inclination β through point 0,

$$\text{Slope} = \tan \beta = \frac{c}{q} = \frac{kq}{q} = k = \bar{c}$$

We conclude that, when total cost is directly proportional to output, the relation between c and q is represented by a straight

---

site the angle to the length of the shorter of the two sides adjacent to the angle. The student can compare this definition with the right-triangle figure OBP in Chart 4. As a special case, the tangent of a zero angle is zero. The definition, with due care for signs, can be extended to angles that are not acute. If, for example, the angle between OQ and OA is 120°, the tangent of the angle is the ratio obtained by dividing the length (positive) of a vertical line—measured from OQ, as extended leftward, to A—by the horizontal length (negative) from O leftward to the foot of that vertical line.

[1] These compound symbols will appear, in various forms, in our further work. We must form the habit of remembering that they are mere abbreviations: tan α means "tangent of α," just as log x means "the logarithm of x," and $\sqrt{2}$ means "the positive number whose square is 2."

line through (0, 0) having a slope equal to the average cost, which is constant.[1]

Instead of a description of the cost situation in terms used in the first assumption, that total cost is directly proportional to output, the description might equally well have read "average cost is constant." This alternative verbal description of the relation between cost and output would lead to precisely the same symbolic representation as above. Thus, using the definition of average cost

$$\bar{c} = \frac{c}{q}$$

and observing that our assumption now states that $\bar{c}$ is a constant, which we shall call k, though we might use any other letter that we agreed would represent a constant, we have

$$\bar{c} = k \tag{4}$$

and

$$\frac{c}{q} = k$$

and, hence, solving for c explicitly

$$c = kq \tag{1}$$

as before. This treatment illustrates the general principle that alternative verbal assumptions, so long as they are strictly equivalent, yield the same symbolic relation between the variables. Conversely, and this frequently leads to confusions in our thinking, though it should not do so, a single symbolic formula may admit of several alternative, though strictly equivalent, verbal interpretations.

However, the second form of the first assumption, that average cost is constant, need not lead to the symbolic formula given in Equation (1). It does so, and then unavoidably, only if we insist on a symbolic representation in terms of q and c. If we were content with a representation in terms of the variables

[1] This presupposes the scales chosen as specified: the same distance represents a unit of q as a unit of c. A serious, and in no sense helpful, modification of the conclusion would otherwise be needed.

q and $\bar{c}$, *i.e.*, with an equation showing the relation of $\bar{c}$ to q, the new form of the assumption immediately yields

$$\bar{c} = k \tag{4}$$

where k is the letter chosen to represent the constant. The variables in this case are q and $\bar{c}$, not q and c. It so happens that q does not appear in the equation; but this does not mean that Equation (4) shows no relation between q and $\bar{c}$, it merely means that the relation is of an extraordinarily simple sort, *viz.*, that $\bar{c}$ does not change as q changes. Equation (4), granting our definition of average cost, is just as good a representation of the cost situation of the specified producer as Equation (1). We saw just above that Equation (1) can be obtained from Equation (4) and on page 10 that (4) can be obtained from (1).

This interchangeable quality of the two equations is the symbolic counterpart of the fact that the two verbal forms of the assumption are strictly equivalent. The matter is of little practical importance in such a simple problem as that here under study, but it will appear below that very great advantages frequently arise from starting out with a properly chosen form of verbal statement and the equation related thereto. In many such problems, we must start out with some one form of verbal statement; we then use that as a basis of setting up the corresponding equation, pass from that equation to another corresponding or equivalent equation, and only then discover an alternative verbal form of the original statement. In essence, this is one of the great achievements of the mathematical process; it enables us, by symbolic operations, to discover unknown verbal relations equivalent to or implied by a given verbal relation.

We can, of course, plot a chart of Equation (5) which is the specific case of Equation (4) when k is 3

$$\bar{c} = 3 \tag{5}$$

To plot, axes of measurement are again chosen; the horizontal axis will again be taken for measuring q, but now the vertical axis will be used for measuring $\bar{c}$ rather than c. Scales will need to be chosen again; though the old scale of Chart 3 can be retained for q, the scale for $\bar{c}$ will differ from that for c. Equation (5) has a locus that is a straight line and the line can be located

by plotting two of its points such as (0, 3) and (4, 3).[1] The result appears in Chart 5. The line is horizontal, at a distance 3 above OQ.

This chart is a graphic representation of our assumption, and just as good a representation as Chart 3. The two representations differ from each other because they show the relations

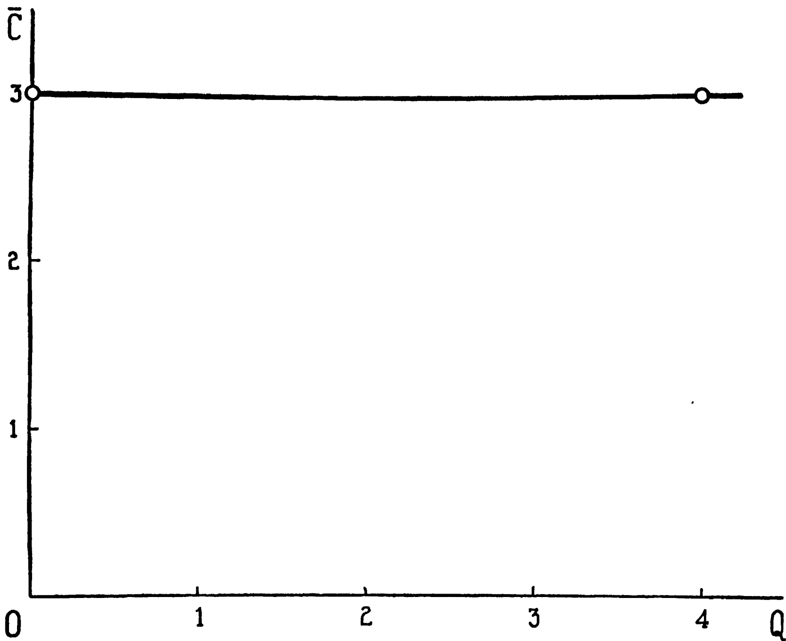

CHART 5.—Fixed average-cost function (proportional total cost).

between different pairs of variables, Chart 3 between q and c, Chart 5 between q and $\bar{c}$. There is no reason to prefer one representation to the other unless we happen to be more interested in studying the relation to q of one variable, *e.g.*, c, than of the other. Either chart is a completely satisfactory graphic representation of the cost situation for the specified producer. Once we have either chart, we do not need the other, but having both

[1] As before, students familiar with analytic geometry will know that Equation (5) is represented by a straight line and that the line is horizontal. We give no proof of this, but other students may reduce their doubts by the kind of testing suggested in footnote 2, p. 10.

charts may be exceedingly helpful to an understanding of the facts.[1]

When two charts are associated as are Charts 3 and 5, through the fact that one of them is derivable from the other—in this case, Chart 5 is derivable from Chart 3 because, for each q, $\bar{c}$ is derivable from c by a definite rule implied in the definition of average cost—and thus merely represents the other in a different form, one of the pair of charts is called the *image* of the other. In some problems the rule of passing from a chart to its image may be much less simple than that used here; here the rule is merely that the $\bar{c}$ of Chart 5 is the c of Chart 3 divided by q. We shall find below numerous instances of image charts, and their contribution to an understanding of the facts under study will often assume high importance.[2]

The cost situation indicated by the assumption under examination in the foregoing analysis—the assumption (1) that total cost is directly proportional to output, or (2) that average cost is constant—is generally described in economics by the term *constant cost.* Manifestly, this means constant average cost; the term might advantageously have included the adjective "average," or the qualifying phrase "per unit." Once the accepted significance of the term constant cost is known, the term should thereafter be used in economics as thus defined: Constant cost means constant average cost. Its mathematical representation appears graphically in Chart 5, symbolically in Equation (5), or more generally in Equation (4). Because of the image relationship, an alternative representation is that of Chart 3 or Equation (2) and more generally Equation (1). Which form of representation is to be preferred depends upon the immediate object of inquiry, as will appear in later discussions.

[1] In a problem less simple than that under study, this possibility of getting several supplementary views by using different graphic representations, a first chart and its image or several images, is of great practical importance. It is like having photographs from various angles of a single complicated landscape.

[2] The student of elementary statistics is familiar with the ogive, or cumulative frequency graph. The ogive is manifestly merely an image of the ordinary frequency polygon. The rule of passing from the polygon to the ogive is the rule that successive frequencies be cumulated.

CHAPTER II

GRAPHIC ANALYSIS: CURVES AND EQUATIONS

In Chap. I, the graphic method was applied to the simplest case—an assumed cost situation for which both total cost and average cost are represented by straight lines. The present chapter extends the method to cases in which curves arise instead of straight lines.

Fixed Total Cost. Not because of its practical economic significance, but because its study will aid greatly in certain later stages of our work, we now examine a second and different assumption concerning the relation of cost and output. Assume now that the specified producer has a fixed—we avoid the word "constant" to minimize danger of confusion with the preceding case, called "constant cost" in economic theory—total cost no matter how great his output.[1] Real economic life, nearly all its productive operations involving the application of certain productive factors that tend to vary with the rate of output, does not readily yield instances of this sort. We can, however, visualize a not entirely unrealistic case. Suppose that a workman can, on any and every day he wishes, readily secure employment at his trade at $7 per day. He chooses not to take such employment, but to engage in door-to-door selling, on commission, of a book on the history of medicine. His output q is the number of units, books, sold per day; his total cost c is fixed at $7 per day.[2] This kind of cost, the doing without revenue from one operation

[1] Strictly, this states the case with unnecessary rigidity; instead of saying "no matter how great his output," we might have said "for all amounts of output within the range under study." This is a point that we neglect here, as we consider total cost fixed for all values of q; but the point will be raised again later, where it has practical importance (p. 32).

[2] We waive here any costs, or savings in cost, involved in differences in attractiveness or irksomeness of the two types of work. In all discussion of cost in this book, cost is taken in the sense of the pecuniary expense of production and not in the sense of those intangible sacrifices that often influence an individual's appraisal of the "cost" of doing some particular act or securing some particular benefit.

in order to undertake a different operation, is a simple and imperfect illustration of what economic theory calls *opportunity cost.*

The symbolic representation of this cost situation is

$$c = f = 7 \tag{6}$$

where f is a constant, fixed at \$7. Both c and f are, of course, measured in units (dollars) per day. Chart 6, showing points

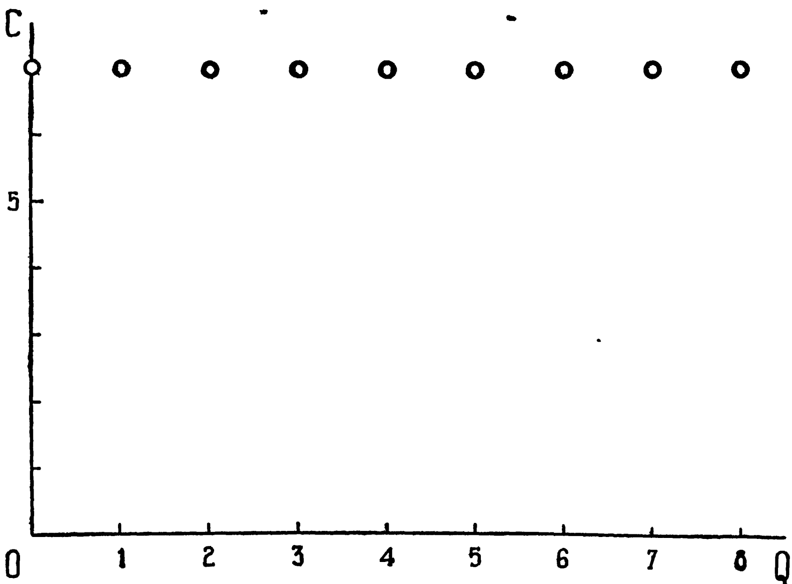

CHART 6.—Fixed total cost, discontinuous case.

along a horizontal line, is the graphic representation of the producer's cost for various q.

How about his average cost? Using the general definition of average cost given by

$$\bar{c} = \frac{c}{q} \tag{3}$$

we get from (6)

$$\bar{c} = \frac{f}{q} = \frac{7}{q} \tag{7}$$

To secure a graphic representation of Equation (7), we need pairs of values of q and $\bar{c}$; such pairs, for selected whole-number (called

integral) values of q, appear in Table 2. The table includes no entry for q as zero: if he sells no book, his total cost for the day is still \$7, but his average cost must not be stated as $7 \div 0$.[1] We observe further that the table includes no fractional values of q; the case under discussion obviously excludes fractions, because no part of a book can be sold separately.[2]

TABLE 2.—VALUES OF $\bar{c}$ ASSOCIATED WITH SELECTED VALUES OF q, FOR EQUATION (7)*

q	$\bar{c}$	q	$\bar{c}$
1	7	4	1.75
2	3.5	5	1.4
3	2.33	6	1.17

* Units: for q, 1 book; for $\bar{c}$, dollars per book.

Chart 7 shows points plotted to represent the several pairs of values given in Table 2. These points are not joined by a curve or by successive line segments because, as noted above, fractional values of q have no significance in this problem. The chart, however, inevitably suggests such a curve to a mind at all accustomed to graphic statistics; visualizing such an imaginary curve is a perhaps unconscious step in our reaching the two important inferences from the chart. Those inferences are (1) the points drift downward toward the right, and (2) this drift becomes increasingly gradual as our eyes pass to the right.[3] The fact that the points do not lie upon a horizontal line, like those of Chart 6, means that, although total cost is constant, average cost is not constant. This fact is also revealed symbolically by

[1] Division by zero is not a permissible mathematical operation. One sometimes acquires the mistaken notion in elementary algebra that such a ratio as $b \div 0$ has a meaning and indeed, if b is some constant other than 0, "equals infinity." We merely assert now that this is wrong, that the noun "infinity" is an abomination to all who wish to think carefully in mathematical terms; and we must reserve further discussion until the doctrine of limits is developed (p. 41).

[2] This is the sort of situation that yields what elementary statistics calls a *discrete* series. For example, if the workman sold books, over a period of many days, we could tabulate a frequency series showing the number of days in which he sold no book, one book, two books, etc. Such a series would be discrete: fractional values of the variate q would be without meaning.

[3] This second observation is equivalent to saying that the curvature of the curve, if in fact the curve were drawn, is of the concave-upward type.

Equation (7): the appearance of q in the right side of the equation forces $\bar{c}$ to vary as q varies.

In order to get rid of the limitation, imposed by the fact that the unit of sale (output) is indivisible, and thus be able to dispense with the treatment of our plotted points as isolated

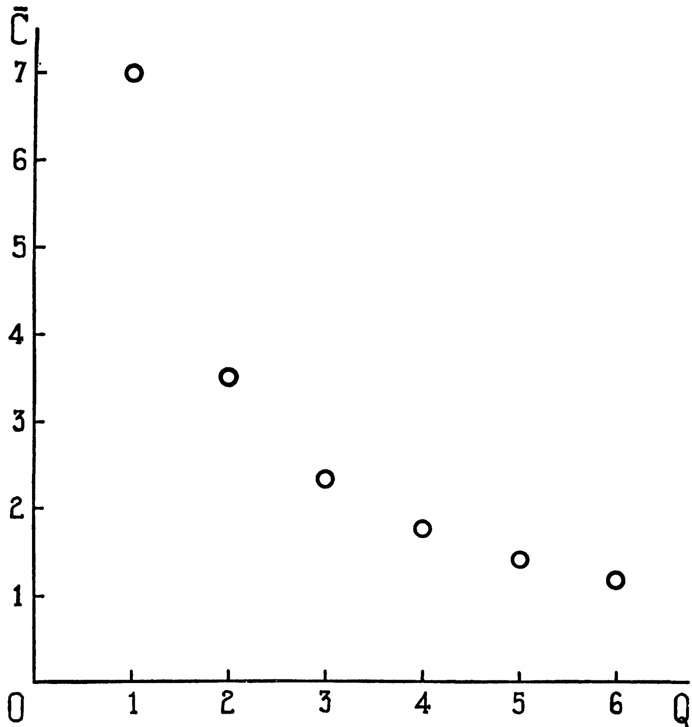

CHART 7.—Average cost for fixed total cost, discontinuous case.

or discrete, consider a modification of the foregoing illustration. Suppose the salesman is selling fire insurance which, though he records his sales in a unit of $1,000 of insurance, can (we suppose) be sold in any amount. Here the unit of output is divisible, and we are warranted in considering fractional values for q.[1] Equa-

[1] In actuality, of course, perfect divisibility is not possible—all conceivable fractions cannot be considered. Thus, the smallest unit in this case might be $10; the rules of the company might prevent sales in any amounts except multiples of 10. In every actual economic problem, some such ultimate limitation on divisibility is imposed by practical or technical considerations. But where, as in the case now supposed, the reported unit can be divided into tolerably small fractions, we treat the variable q as truly continuous, *i.e.*, perfectly divisible.

tion (7) still holds, q now being expressed in the unit \$1,000. To plot the equation, we calculate the pairs of values in Table 3. The corresponding points—except (0.01, 700) and (0.1, 70), which are so high that convenience of scale dictated their omission from the chart—are shown in Chart 8. Here a continuous curve has been drawn through the plotted points, because we feel justified in regarding q as capable of taking on any value, fractional to any degree of fineness.

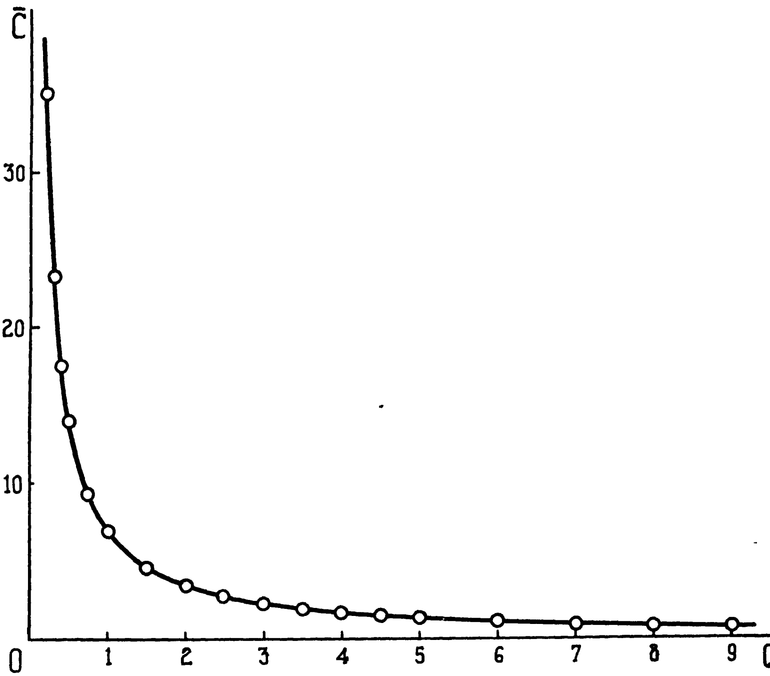

CHART 8.—Average cost for fixed total cost, continuous case.

Examination of the chart, or the table, shows that the plotted points are much closer together, in the sense of lateral (horizontal) distance, at the left than at the right—the fractional values are taken finer and finer as we pass to the left. This is because a preliminary testing of integral (whole-number) values for q showed that the curve changes more rapidly as q changes for small values than for large values of q, *i.e.*, at the left than at the right.[1]

[1] There is no trustworthy rule to guide us in deciding what fractions to substitute, where along the range of variation to substitute fine fractions,

The curve of Chart 8 is a complete graphic representation of Equation (7) and leads again to the two important inferences reported above in examining Chart 7.

TABLE 3.—VALUES OF $\bar{c}$ ASSOCIATED WITH SELECTED VALUES OF q FOR EQUATION (7)*

q	$\bar{c}$	q	$\bar{c}$
0.01	700	2.5	2.8
0.1	70	3	2.33
0.2	35	3.5	2
0.3	23.3	4	1.75
0.4	17.5	4.5	1.56
0.5	14	5	1.4
0.75	9.33	6	1.17
1	7	7	1
1.5	4.67	8	0.88
2	3.5	9	0.78

* Units: for q, \$1,000; for $\bar{c}$, dollars (of cost) per \$1,000 (of sales).

Composite Total Cost, First Example. We pass now to a third assumption governing the cost of our producer: His total cost is a composite of two parts, one fixed and the other directly proportional to output.[1] His composite total cost c is thus made up of two parts c_f and c_p, where the subscripts f and p distinguish the fixed and variable parts.[2] From Equations (1) and (6)

$$c_p = kq \qquad \text{and} \qquad c_f = h$$

and how fine, and where to substitute coarser fractions, and how coarse. Experience teaches, and we presently learn to calculate enough pairs of values, in various portions of the range, to enable us to draw the desired curve as precisely as we wish. In any particular case, some experimentation may be necessary; and we may be obliged to calculate and plot some additional, and intermediate, points when those already plotted appear inadequate to guide our drawing of the curve. In some parts of the curve, it may not be necessary to plot for all integral values of q.

[1] This assumption, unlike those which have gone before, more nearly represents actual situations, of a very simple type, in economic life. Even here, however, careful economic reasoning brings out questions as to the likelihood of such a law of cost prevailing for any considerable change in q.

[2] We use here "total cost" to mean the entire cost for output q and the term "composite" to indicate that it is made up of parts. Economics sometimes uses total in the sense in which we use composite, and the usage specified here should be kept in mind as we proceed.

where h corresponds to the f of Equation (6). Hence, as

$$c = c_p + c_f$$

the symbolic formula for this new assumption is

$$c = kq + h \tag{8}$$

where k and h are constants. The units for q and c are supposedly known: let them be 100 tons and \$1,000.

In order to render this general formula specific, definite numerical values are assigned to k and h; although we might choose any numbers for this purpose, it is convenient for illustration to select

$$k = 3, \qquad h = 16$$

and the equation becomes

$$c = 3q + 16 \tag{9}$$

This is the equation of a straight line, and it can therefore be plotted by locating two of its points.[1] One such point, easily calculated and plotted, is for q zero, for which c is 16.[2] A second convenient point is (4, 28). These two points are plotted in Chart 9; as we regard q capable of continuous variation and

[1] That the equation gives a straight line will be known to those familiar with elementary analytic geometry. How is this known? It is known by the facts that each variable q and c enters the equation only to the first power and there is no term of the equation involving the product q times c. In technical terminology, such an equation is of the *first degree*, or *linear*, in q and c. Again we omit proof that the locus of Equation (9) is a straight line, but suggest that the student test it in the manner indicated in footnote 2, p. 10.

[2] Similarly, taking c zero gives q as $-1\frac{6}{3}$. These values of c and q, obtained by taking q and c, respectively, zero, are called the *intercepts* of the line: they give the distance from O at which the line cuts the axes OQ and OC. The c intercept is 16, and the q intercept is $-1\frac{6}{3}$. Plotting these two points completely determines the straight line. As, however, the q intercept is negative and the line therefore cuts the OQ axis on the left of O, and as economic considerations imply that the chart should not be drawn left of O, we do not use this second intercept $q = -1\frac{6}{3}$ in locating the line. Instead, as shown above in the text, we take some other point, a point for which q is positive. Nevertheless, if we had extended the OQ axis to the left and plotted the q intercept as well as the c intercept, we should have found precisely the same line as that determined by (0, 16) and (4, 28). There is only one line representing Equation (9) no matter what points are used in plotting it.

therefore of taking on fractional values as well as integral values, a continuous line is drawn through the two points. This line is the graphic representation of the costs of the producer.

The line shows that (1) the lowest cost (ignoring the negative portion of the chart, purposely not plotted) is 16 and appears for q zero, (2) the cost increases as we pass to the right. Both of these facts could be inferred from Equation (9), fact 2 by noting that q is multiplied by a positive constant, and fact 1 as a necessary corollary thereto. The chart is therefore merely a different form of representation, leading to no inferences not derivable from the equation.

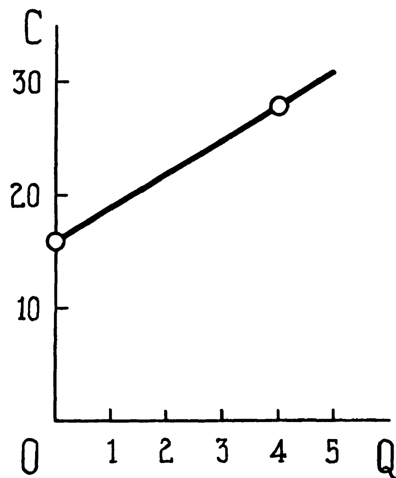

CHART 9.—Total cost a composite of a fixed and a proportional element.

The location of this line could be described otherwise than by specifying the two points (0, 16) and (4, 28), *e.g.*, by specifying (0, 16) and the inclination of the line, or by specifying (0, 16) and the slope of the line. The last of these possibilities is worth examining for the light it throws on a problem treated below. Chart 10 reproduces the line as in Chart 9, but here a supplementary horizontal line has been ruled through B, (0, 16). L is any point (q,c) on the line BA, but cannot be (0, 16). The angle α, between BH and BA, is the inclination of the line BA, and tan α is the slope of BA.[1] Using the definition of tan α (page 13)

$$\tan \alpha = \frac{ML}{BM}$$

[1] In order that the present analysis shall be valid, the scales must be chosen so that a unit of q, 100 tons, is measured by the same distance as a unit of c, $1,000. This was not essential for the purpose of Chart 9, and there the scales were not chosen with units of q and c equal. In elementary statistics we frequently use scales that do not represent the two units, one for each variable, by an identical distance. An illustration of this appears in Chart 1. In order, however, that the slope of the line shall work out as in the text herewith, the scales must represent the two units equally, as in Chart 10.

But, as

$$BM = ON = q$$

and

$$ML = NL - NM = NL - OB = c - 16$$

we have

$$\tan \alpha = \frac{c - 16}{q}$$

and, as by Equation (9)

$$c = 3q + 16$$

$$\tan \alpha = \frac{3q + 16 - 16}{q} = 3$$

Hence the slope of the line BA is 3, and this is manifestly the multiplier (customarily called the *coefficient*) of q in Equation 9.[1]

We turn now to a study of average cost for this producer. By definition of average cost

$$\bar{c} = \frac{c}{q}$$

and, from Equation (8)

$$\bar{c} = \frac{kq + h}{q} = k + \frac{h}{q} \tag{10}$$

and, in the specific case, from Equation (9)

$$\bar{c} = \frac{3q + 16}{q} = 3 + \frac{16}{q} \tag{11}$$

[1] This is a general fact. If we start with

$$c = kq + h \tag{8}$$

a similar analysis would show that for this general line (its inclination being represented by β)

$$\begin{aligned} \text{Slope} = \tan \beta &= \frac{c - (\text{the } c \text{ intercept})}{q} \\ &= \frac{c - h}{q} \\ &= \frac{kq + h - h}{q} \\ &= k \end{aligned}$$

For any equation of a straight line, giving the dependent variable explicitly as an algebraic function of the independent variable, the slope of the line (plotted always with the units for both variables taken equal) is the multiplier (generally called "coefficient") of the independent variable.

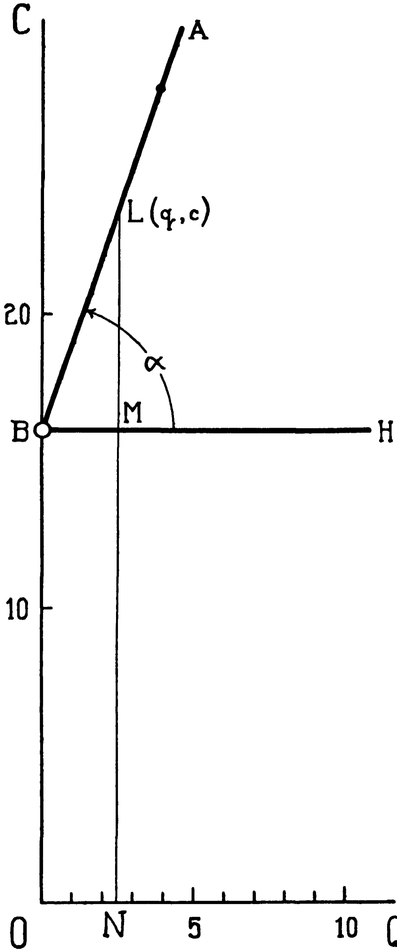

CHART 10.—Inclination of line of composite cost.

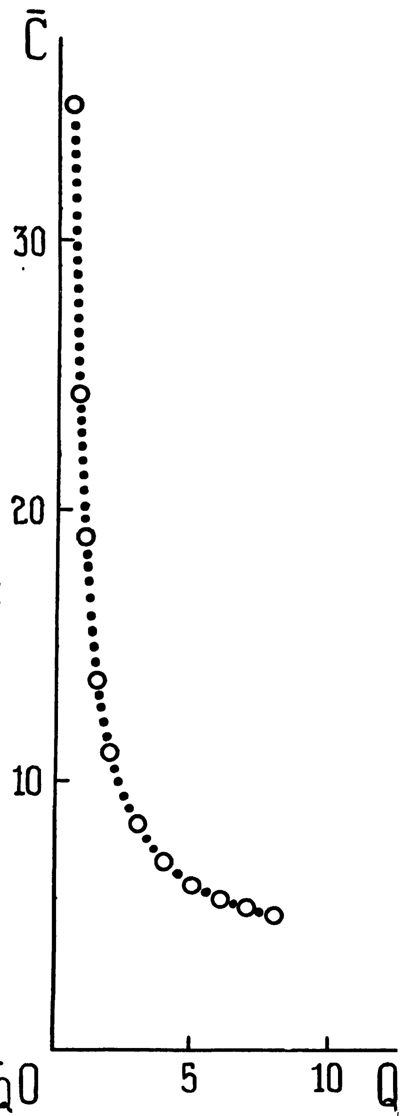

CHART 11.—Average cost for composite total cost shown in Chart 10.

To plot Equation (11), pairs of values of q and $\bar{c}$ are needed, and they appear in Table 4. The points are plotted in Chart 11, and as q is regarded capable of taking on all values, a curve is drawn through the points.[1]

TABLE 4.—VALUES OF $\bar{c}$ ASSOCIATED WITH SELECTED VALUES OF q, FOR EQUATION (11)*

q	$\bar{c}$	q	$\bar{c}$
0.01	1,603	1.5	13.7
0.1	163	2	11
0.2	83	3	8.33
0.3	56.3	4	7
0.4	43	5	6.2
0.5	35	6	5.67
0.75	24.3	7	5.29
1	19	8	5

* Units: for q, 100 tons; for $\bar{c}$, $1,000 per 100 tons.

The chart is the image of Chart 10. Chart 11 shows that (1) the average cost declines as we pass to the right, as q increases, and (2) the rate of this decline becomes more gradual as we pass to the right. These are precisely the same inferences that we drew for the case of a producer with constant total cost (page 20); whereas, for a producer with total cost directly proportional to output, we found (page 16) that the average cost curve was a horizontal line the elevation of which above the OQ axis was equal to the proportionality factor [3 in the illustration on page 16, or k in the general case of Equation (8)].

We can, in fact, regard the $\bar{c}$ of Equation (11) as made up of two parts. First is the constant part 3, due to the portion of total cost that is directly proportional to output. Second is the variable part (variable because it depends on q) $16/q$, due to the portion of total cost that is fixed at 16. These two parts could be plotted separately, a table of pairs of values being needed for the

[1] For convenience of scale, the first five pairs of values in Table 4 are not plotted. Note that no entry appears in the table for q zero; this would involve division by zero, about which see footnote 1, p. 20. In entering fractional values of q in the table, fractions are taken closer together for that part of the curve which changes direction more rapidly, see footnote 1, p. 22.

second part, as it is variable; the two portions are shown, except for the high points at the left, by the dashed line and dotted curve of Chart 12. To get the solid curve of the chart, which is exactly the same as that of Chart 11, the two portions are added; for every value of q such as ON, the two ordinates NL and NM are measured and added to get NH. This is precisely equivalent, subject to accuracy of measurements, to the additions actually performed in getting the $\bar{c}$ of Table 4. Thus, graphically in Chart 12, as well as symbolically in Equation (11), average cost for this producer appears manifestly as composed of two parts. One is constant (horizontal straight line) and due to that part of total cost which is directly proportional to output; the other is variable (the declining curve, concave upward) and due to the fixed element in total cost.

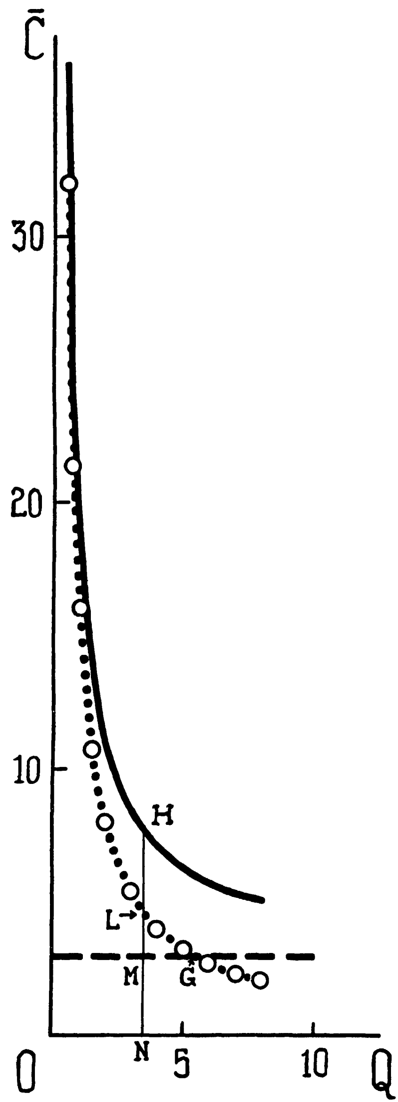

CHART 12.—Separate elements in composite average cost of Chart 11.

Before passing to a fourth assumption, we can helpfully dwell briefly upon some important economic implications of the third assumption. The foregoing analysis shows that, whenever total cost is a composite of two parts, a fixed part and a part directly proportional to output, average cost is also a composite of two parts.[1] That part of average cost

[1] Again the distinction between "composite" and "total" should be

corresponding to the fixed element in total cost is variable, and its variation inevitably takes the form found above—average cost declines as output increases. That part of average cost corresponding to the variable (proportional) element in total cost is constant, and its amount is inevitably equal to the proportionality factor [the coefficient of q in such an equation as (8)].

Manifestly there is some value of q for which the two parts of average cost are equal. This value, in the specific case represented by Chart 12, pertains to the point G where the dotted curve (the average-cost image of the fixed element in total cost) cuts the dashed line (the average-cost image of the proportional element in total cost). In the specific numerical illustration represented by Chart 12, the q for G is $16/3$; in the general case of Equation (10), it is h/k. To the left of G, the dotted curve is above the dashed line; to the right of G, the reverse is true. In other words, to the left of G, average cost due to the fixed element in total cost exceeds the average cost due to the proportional element in total cost; to the right of G, the reverse is true.

The portion of cost that depends upon output, the variable portion of total cost, is called *prime* cost; the portion of cost that is fixed, which is thus independent of output, is called *supplementary* cost, or sometimes *overhead* cost.[1] The average cost incident to supplementary (overhead) cost takes the shape, although the general direction and degree of curvature may be different, shown by the dotted curve of Chart 12. The average cost incident to prime cost may, however, assume a more complicated form than the dashed straight line of Chart 12; its form depends upon the way in which prime cost depends upon output. Only if prime cost depends upon output according to the simple, and rather unrealistic, rule of direct proportionality specified by our third assumption does the corresponding average cost appear as a horizontal straight line.

---

noted: the former refers to mere combining of the two elements of cost, the latter refers to the aggregate cost for output q as distinct from the average cost.

[1] These terms appear to have very simple definitions in this illustration; but, for cost situations more realistic and less simple than that specified by our third assumption, the significance of the terms becomes increasingly intricate, and often appears obscure to the student. The present illustration merely indicates the distinction, in its simplest terms, between prime and supplementary costs.

Finally we infer from Chart 12, since the solid curve representing composite average cost appears to get closer and closer to the horizontal dashed line as we pass to the right but can never drop below that line, that the producer can, by making his output larger and larger, bring his average cost down nearer and nearer to a fixed limiting level [3 for Chart 12, k for the general case of Equation (10)]. This is the first suggestion of the notion of a *limit,* to which more specific attention will be given in the next chapter.

Composite Total Cost, Second Example. Now let the cost situation of the specified producer be subject to a fourth, and still more complicated, assumption. Suppose that his total cost is made up of two parts c_f and c_v, where the subscripts f and v indicate fixed and variable, and that these parts are represented symbolically by

$$c_f = f$$
$$c_v = aq^2 + bq$$

where a, b, and f are constants. As

$$c = c_v + c_f$$

the general formula for this case is

$$c = aq^2 + bq + f \tag{12}$$

The reader may inquire how the formula for c_v, in terms of q, is known for the supposed producer. We are assuming that the producer, although he might have used the empirical method, finds his formula by the a priori method, that he works out the formula from known facts about his system of production and the attendant costs. Thus, presumably he knows that part of his total cost, represented by f, is fixed regardless of output; part, represented by bq, is directly proportional to output; and the rest, represented by aq^2, is directly proportional to the square of output.

Just how, from a technical or organizational point of view, an element of cost of the third sort, proportional to the square of the output, can arise in practical economic life may at first baffle the student. Without attempting to catalogue possible cases, we suggest only one; this may remove the student's doubt that a cost of such type could exist. Suppose that, for a limited range

of output, the selling cost per unit of a patented article just being introduced into consumption is proportional to the volume sold; in order to push sales of the novelty to an extent securing some volume q, the advertising and other sales expenditure per unit sold is some constant a times the volume q. The total

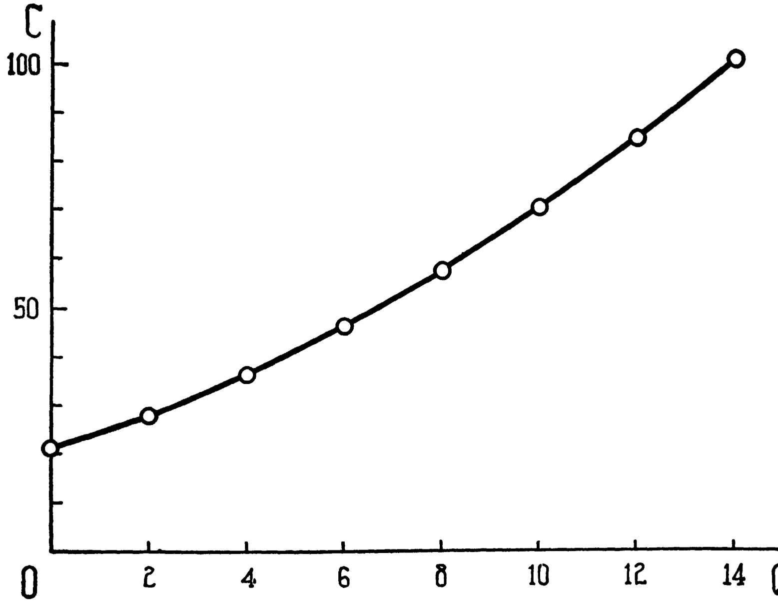

CHART 13.—Composite total cost with three elements: fixed, proportional, and dependent on second power of output.

selling cost for a units would then be[1] aq^2. If total manufacturing cost takes the form indicated by the assumption in the preceding section, *viz.*, $bq + f$, total cost, including selling and manufacturing costs, would be as given by Equation (12). Output, *i.e.*, production, here includes selling as well as manufacturing operations.[2]

[1] Such a formula for sales cost would, as suggested above, presumably apply only for a limited range of possible values of q. If the objective were chosen outside this range, a much smaller, or a much larger, number of units in the early phase of popular acceptance of the novelty, a different formula for sales cost might apply. We are concerned only with some q within this supposed range.

[2] The student will understand, of course, that in modern industry such a term as aq^2 might occur in the cost formula for manufacturing alone. In fact, economic theory suggests that a third element, involving q^3, may enter, and perhaps other elements.

To render the case specific, we assume a, b, and f are $\frac{1}{5}$, 3 and 21, respectively. The equation becomes

$$c = \tfrac{1}{5}q^2 + 3q + 21 \tag{13}$$

The corresponding points, Chart 13, are plotted from the pairs (with some omissions) of values shown in Table 5, and the plotted points are joined by a curve, because we regard q as capable of taking on fractional values.[1]

TABLE 5.—VALUES OF c ASSOCIATED WITH SPECIFIED VALUES OF q, FOR EQUATION (13)*

q	c	q	c
0	21	10	71
2	27.8	12	85.8
4	36.2	13	93.8
6	46.2	14	102.2
8	57.8	15	111

* Units: for q, one dozen articles; for c, \$100.

The curve shows that (1) some cost appears even if output is zero, (2) total cost rises as output increases, and (3) total cost rises at an increasing rate as output increases. Inference 3 is implied by the fact that the curve is concave upward.[2]

Before proceeding to the corresponding average cost analysis, a more detailed graphic study of Equation (13) is informing. For this purpose, we use the scheme adopted in Chart 12 for plotting separately the two terms of Equation (11).[3] As the right

[1] For the use to be made of this chart, the scales need not be chosen so that both units are represented by the same distance. The curve is plotted only between q equal to zero and q equal to 14, because we are assuming that the formula for cost, particularly for sales cost, holds only in that range. So far as the mathematical properties of Equation (13) are concerned, however, we could prolong the curve beyond the limits of this range. The physically possible fractions, of course, are multiples of $\frac{1}{12}$ of the unit used for q in the chart.

[2] Inference 3, as stated, involves the word *rate*. The rate concept will be developed more precisely (p. 81) and will be shown to be associated with the curvature of the curve.

[3] A similar attack might have been made upon Equation (9), but was omitted. Thus, a chart corresponding to Chart 9, but showing separately also the curves due to the two parts $3q$ and 16, would have been found. Each of these separate curves would have been straight lines, the first inclined, the second horizontal.

side of Equation (13) contains three terms, we might plot each one separately, securing such a figure as Chart 14. The constant term 21 yields a horizontal straight line, the proportional term $3q$ yields an inclined straight line, and the second-degree term $\frac{1}{5}q^2$ yields a concave-upward curve.[1] The solid curve of the chart, the same curve as that of Chart 13, could be plotted by adding, for each value of q, the three ordinates of the three elements, *viz.*, the dotted curve and the two straight lines (see page 29).

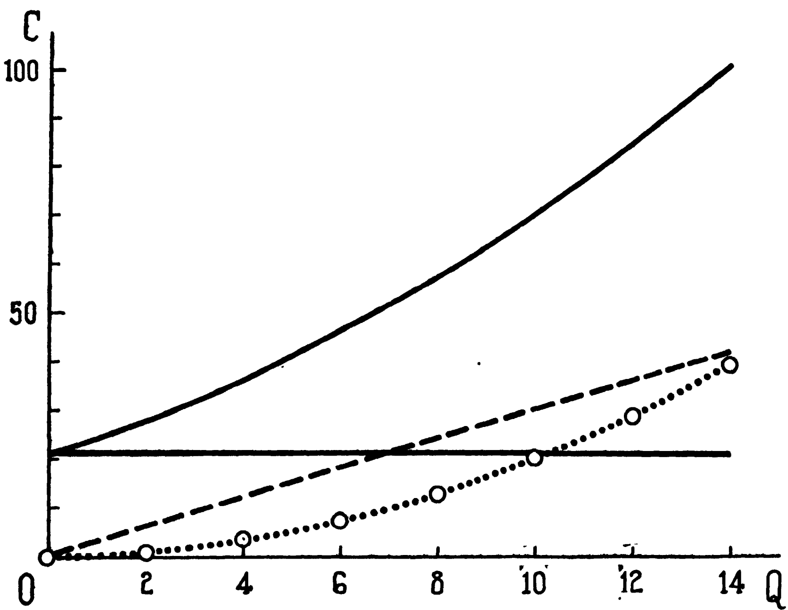

CHART 14.—Separate elements of composite cost shown in Chart 13.

Another possibility consists in breaking the right side of Equation (13) in two parts, *e.g.*, the sales cost $\frac{1}{5}q^2$ and the manufacturing cost $3q + 21$. This yields the dotted curve (the same curve as that dotted in Chart 14) and the inclined line of Chart 15. Otherwise, the right member of Equation (13) can be split in two parts, fixed cost 21 and variable cost $\frac{1}{5}q^2 + 3q$, yielding the horizontal line and dotted curve of Chart 16. In each of these charts—14, 15, and 16—addition of the ordinates of the elements

[1] The student is now familiar with the plotting of straight lines and could draw each of these readily. Plotting of the concave curve, of course, requires a table of values, but this is not reproduced here.

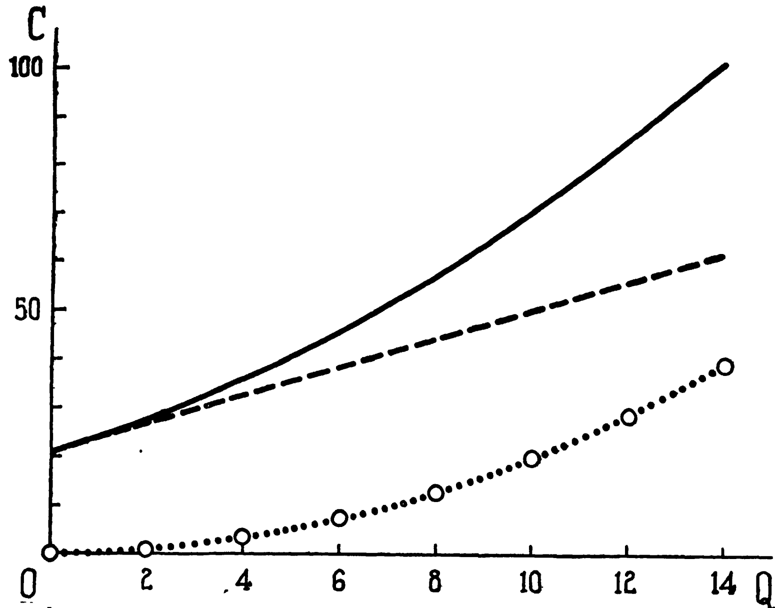

CHART 15.—Composite cost of Chart 13 separated into linear and curved portions.

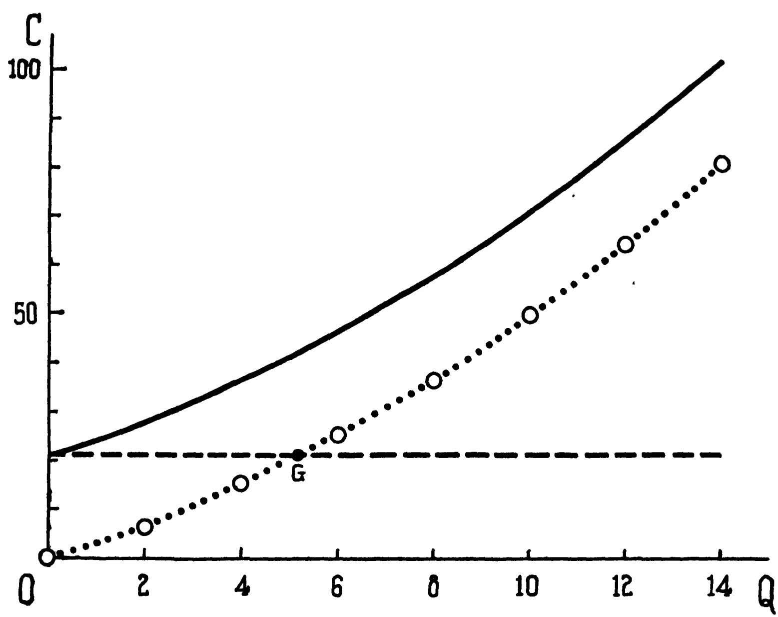

CHART 16.—Composite cost of Chart 13 separated into fixed and variable portions.

(curves or lines), for any value of q, gives the corresponding ordinate of the curve of Chart 13 by precisely the method of addition used in combining the dotted curves and line of Chart 12 to yield the solid curve thereof.

Chart 16 shows fixed cost the same as variable cost at point G, where q is approximately 5.2. Left of G, fixed cost exceeds variable cost; right of G, the reverse is true. Chart 15 shows no corresponding intersection of the sales cost and manufacturing cost curves, because they intersect outside the range, ending for q at 14, to which we have restricted q. Mathematically there is, however, such an intersection; it occurs at some point H (not shown in Chart 15) for which q is approximately 20.2. Left of that point, sales cost is less than manufacturing cost.[1]

The average cost, corresponding to the general formula in Equation (12), is

$$\bar{c} = aq + b + \frac{f}{q} \tag{14}$$

and that for Equation (13) is

$$\bar{c} = \tfrac{1}{5}q + 3 + \frac{21}{q} \tag{15}$$

The curve representing Equation (15) is plotted in Chart 17 by the use of a table, not reproduced, of pairs of values of q and $\bar{c}$. Unlike any of the curves heretofore shown, this curve is inclined downward in part of its range (the left part) and upward in the rest of its range (the right part). At some intermediate point, for which q appears to have the value of about 10, the inclination shifts from downward-to-the-right to upward-to-the-right. (In order to bring out this fact, a light horizontal line has been drawn

[1] The q for G is obtained by setting

$$\tfrac{1}{5}q^2 + 3q = 21$$

and solving for q. Likewise, the q for H is obtained from

$$\tfrac{1}{5}q^2 = 3q + 21$$

In each case, the negative solution for q is discarded; it has no economic significance. The student recalls from elementary algebra that the solution of such an equation as the first of these is secured by throwing it in the form $q^2 + 15q = 105$, adding $(\tfrac{15}{2})^2$ to each side, and then taking the square root of each side. Or, the same result is secured by using a formula designed to give at once the solution for q of an equation of the type $aq^2 + bq + c = 0$, where a, b, and c are constants.

through the lowest point.) This point is a minimum of the curve, and we shall learn (page 110) how its position, the corresponding value of q, can be precisely determined. For the present, approximate location of the point can be obtained by

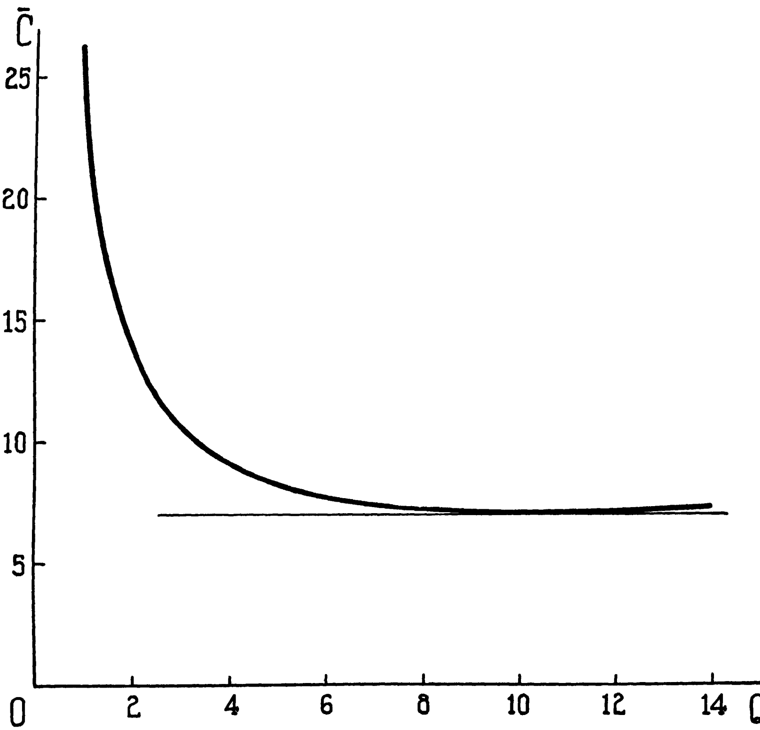

CHART 17.—Average cost for total-cost function shown in Chart 13.

substituting in Equation (15) fractional values of q close to the apparent minimum point revealed by the chart. Thus,

For q:	9.9	10	10.1	10.2	10.3	10.4
$\bar{c}$ is:	7.1212	7.1000	7.0992	7.0988	7.0988	7.0992

where computations for $\bar{c}$ have been carried to four decimals to secure the needed precision. These results suggest that the minimum occurs for q somewhere between 10.2 and 10.3, but no cut-and-try method such as this can give us an exact and unambiguous determination of the minimum.

Again, a graphic representation that shows parts of the right member of Equation (15) separately is informing. For this

purpose, the parts chosen are

$$\tfrac{1}{5}q + 3 \qquad \text{and} \qquad \frac{21}{q}$$

These parts correspond, respectively, to the variable part $\frac{1}{5}q^2 + 3q$ and the fixed part 21 of total cost. The corresponding curves, an upward-inclined straight line and a concave-upward

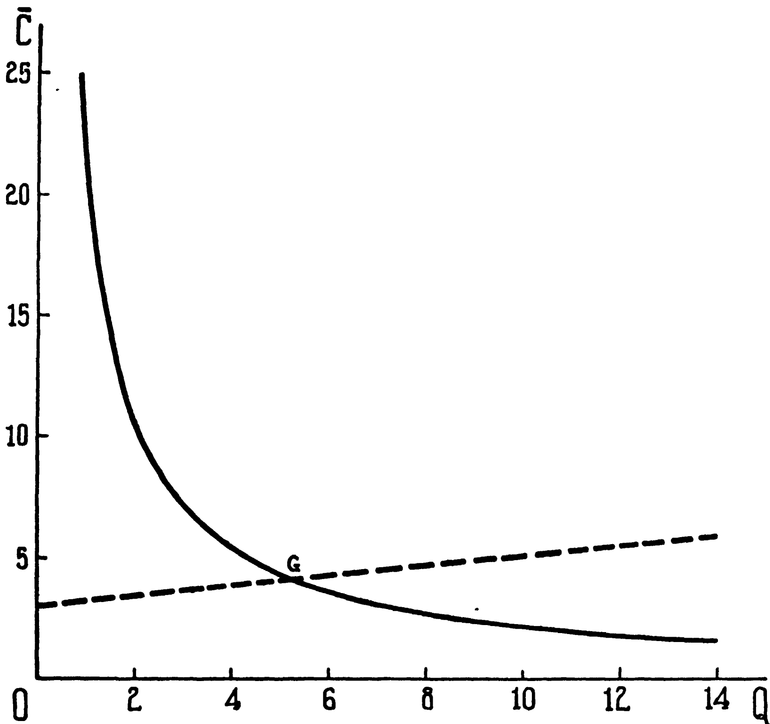

CHART 18.—Portions of average-cost curve of Chart 17 due to fixed and variable portions of Chart 16.

curve, appear in Chart 18.[1] The curve (drawn solid) is inclined downward throughout its length; it has no minimum, in the sense of the minimum of the curve of Chart 17. In truth, such a curve can never, no matter how large q becomes, drop below a particular level; in this case that level is zero. But this is not a minimum in the above sense, it is a limiting value, which

[1] A table of values, not reproduced, is essential for plotting the curve. The student can locate the line by finding two points thereof.

tne ordinate of the curve approaches as it declines, but can never reach.

Addition of the ordinates of the two curves of Chart 18, for any particular value of q, yields the corresponding ordinate of the curve of Chart 17, by exactly the procedure outlined for Chart 12. It now appears that the shift from a downward to an upward inclination, which occurs in the curve of Chart 17 for q about 10.2, is due to the inclined line of Chart 18; the portion of composite average cost due to the curve of Chart 18 never turns upward. In other words, the variable part of total cost, not the fixed part, accounts for the upturn in composite average cost to the right of q equal to 10.2. In still other words, it is the combined effect of the downward inclination of the curve of Chart 18 and the upward inclination of the line thereof which produces the minimum in the curve of Chart 17. That minimum exists because the two constituents of average cost have opposite inclinations.[1]

The economic implication is that whenever total cost is made up of two elements, one fixed (and positive, as in most cost problems of a simple sort) and the other involving, with or without a proportional term, a positive coefficient times the square of the output, average cost must be minimum for some determinable size of output. We remark, however, that the permissible range, for q between 0 and 14 in the above case, might not include the minimum point of the function.

The two curves of Chart 18 intersect at G, for which q is approximately 5.2. At that point, average cost due to the fixed element in total cost exactly equals average cost due to the variable element; to the left of G, the former exceeds the latter; to the right of G, the reverse is true. The q for G is, of course, the same as for the G in Chart 16; this G, like the G of Chart 16,

[1] Instead of Chart 18, another chart could have been provided, showing the right member of Equation (15) split into two different parts, *viz.*,

$$\tfrac{1}{5}q \qquad \text{and} \qquad 3 + \frac{21}{q}$$

corresponding to average sales cost and average manufacturing cost. The only result would have been to shift the straight line down 3 points and the concave curve upward 3 points. The inclination of the line would remain the same as in Chart 18, the shape of the curve would remain as in Chart 18, but the limiting value of the ordinate of the curve would now be 3 instead of zero.

of which Chart 18 is an image, represents the production situation for which fixed total cost equals variable total cost. The corresponding elements of average cost are also necessarily equal.

Before leaving Chart 18, we remark, as should already be seen by the student, that the average cost due to the fixed element in total cost is not fixed but variable, not a horizontal line but a curve. The average cost due to the variable element in total cost is also variable; but its variability, instead of being represented by a curve such as the curve showing variable total cost in Chart 16, is represented by an inclined line. For variable total cost according to a more complicated formula than that of Equation (13) [or, more generally, Equation (12)], the second clause of this inference concerning average cost due to the variable element of total cost would need to be altered as shown below (page 124).

Summary. These first two chapters have developed graphic representation as a means of rendering visual important characteristic facts about several simple types of equations. Many more complicated functional relations can, of course, be represented by charts; although such representation is of sufficient aid in the interpretation of complicated functions to indicate its invariable use by the careful student, many properties of such functions can be revealed only by the more elaborate analytical methods to be developed below. These chapters have also shown the advantage of supplementing one graphic representation by a second, related to the first in some definite way that suggests calling them images of each other. In order to obtain finer tools for the analysis and interpretation of functions that are not exceedingly simple in form, symbolic methods linked with the doctrine of limits are especially helpful. To that doctrine we now turn.

CHAPTER III

LIMITS

Previous chapters have revealed that, even for the simple functional relation involved in constant total cost, such a derived function as average cost is not constant but variable. Although in a case so simple as that of constant total cost a sufficient understanding of the characteristics of both average cost and total cost can be obtained from study of the charts, yet many problems encountered in economic analysis yield functions not readily or adequately explained by simple graphic representations. In many such cases, a derived ratio, of which average cost in the cases already studied is an example, is not only variable, but the effective study of its variation involves the concept of a limit. This concept is of the highest importance as a foundation for analytical methods to be developed later in this book, and the present chapter is concerned entirely with the subject of limits.

We have already encountered the notion of limit in studying (pages 29 and 31) the average cost function

$$\bar{c} = \frac{16}{q}$$

where we saw that, no matter how large q is taken, $\bar{c}$ can never fall as low as zero. To consider this case more generally, suppose

$$\bar{c} = \frac{k}{q} + h$$

where k and h are positive constants. When q is 1, $\bar{c}$ is $k + h$; as q becomes larger, $\bar{c}$ becomes smaller; but, no matter how large q is taken, $\bar{c}$ is always greater than h because the fraction k/q always has a positive value. Under such conditions, h is called the limit of $\bar{c}$, as q becomes infinite. More generally, if for a function $\bar{c}$ and a constant h the difference (without regard to sign)

$$\bar{c} - h$$

can be made smaller and smaller, and less than any previously assigned small number however small, by allowing the inde-

pendent variable q to behave in a prescribed manner, h is called the *limit* of $\bar{c}$ as q behaves in that prescribed manner. In the present instance, the prescribed manner in which q behaves is that it becomes larger and larger. Thus, q can be taken large enough so that $\bar{c} - h$ becomes less than any previously assigned small number.

In the foregoing illustration, the limit was approached as q, the independent variable, became larger and larger. In the bulk of this chapter, however, the functions studied are ratios involving the independent variable, the variable that changes in a prescribed manner, and these ratios approach ascertainable limits as the independent variable becomes smaller and smaller, *i.e.*, as it approaches zero. Here the definition of limit would read: If for a ratio $\Delta y/\Delta x$ and a constant h the difference (without regard to sign)

$$\frac{\Delta y}{\Delta x} - h$$

can be made less than any previously assigned small number, by taking Δx sufficiently small, h is the *limit* of the ratio as Δx approaches zero.[1]

Minimum Average Cost. Suppose that the cost situation of a particular producer is defined, in terms of output and total cost, according to the fourth assumption discussed in Chap. II (page 31), by

$$c = aq^2 + bq + h \tag{12}$$

where a, b, and h are constants. To render the case numerically specific, let the constants be chosen as 0.2, 1.5, and 7. The equation then is

$$c = 0.2q^2 + 1.5q + 7 \tag{16}$$

If we assume that q can take on the value zero and all positive values, including fractions, the curve is as shown in Chart 19. Here the scales have been chosen so that a single distance, horizontally or vertically, represents the units for q and c, because this choice exhibits the geometric and trigonometric relationships in the form desired.[2]

[1] The student should understand that Δx, and likewise Δy, is a compound symbol (see footnote 1, p. 13). It does not mean, and should not be read, "delta times x."

[2] The plotted pairs of values, for q and c, appear in Table 6, page 45.

Choose H as some point (q, c), far out to the right on the curve (reproduced in Chart 20), and, to be specific, let H be (9, 36.7).

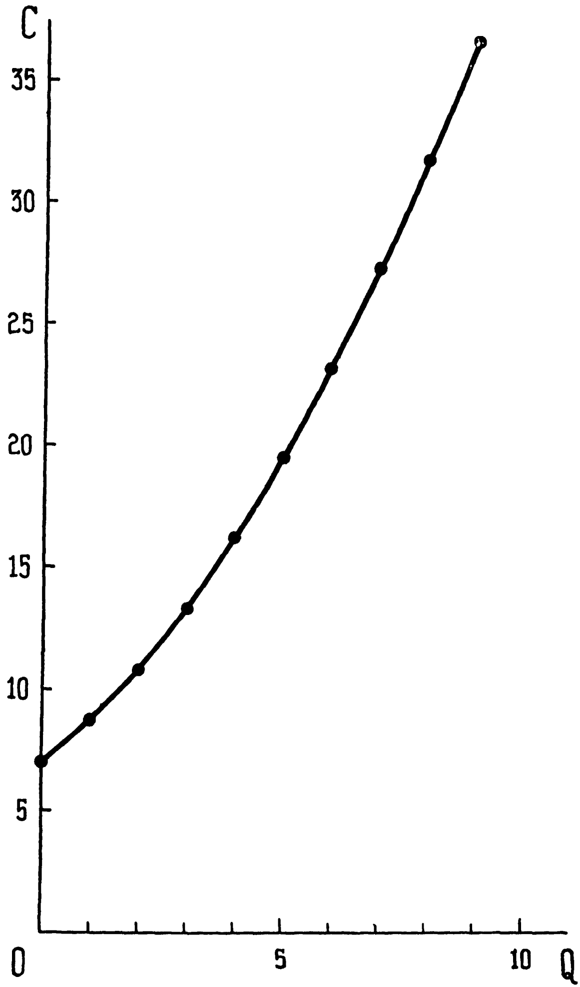

CHART 19.—Composite total cost with three elements, second example.

The average cost for the scale of production indicated by output 9 is

$$\frac{36.7}{9}, \text{ which is 4.078 (approximately)}$$

If the straight line OH is drawn, average cost is manifestly the

slope of the line OH (see page 13). In general, no matter where H, (q, c), is located on the curve, the average cost c/q is the slope of the line OH. Also, if α is the inclination of OH, the average cost is $\tan \alpha$.

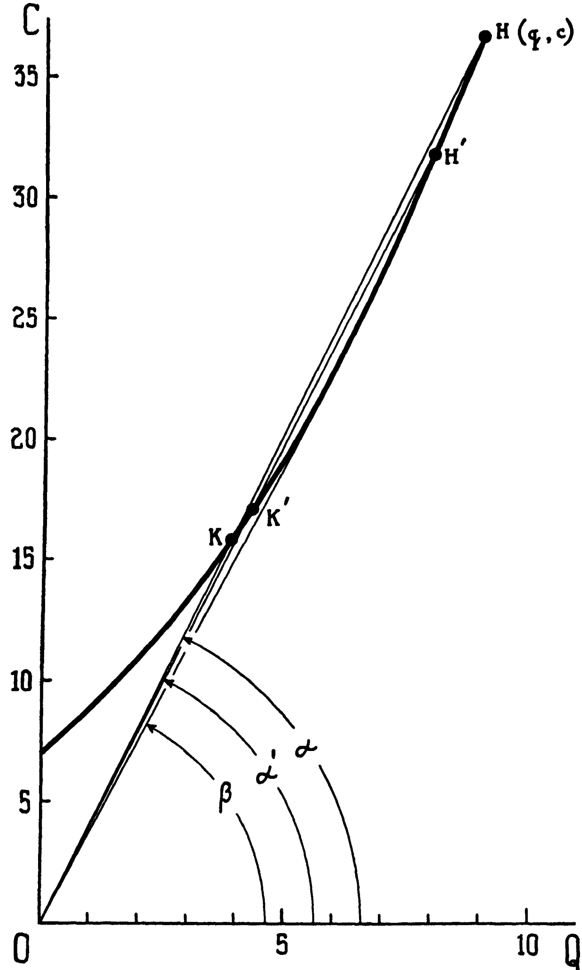

Chart 20.—Inclination of lines reflecting average cost of points of curve of Chart 19.

Referring again to the specific location (9, 36.7) of the point H, we observe that the line OH cuts the curve also in another point K to the left. We infer, of course, that average cost at

TABLE 6.—VALUES OF c ASSOCIATED WITH STATED VALUES OF q, FOR EQUATION (16)*

q	c	q	c
0	7	5	19.5
1	8.7	6	23.2
2	10.8	7	27.3
3	13.3	8	31.8
4	16.2	9	36.7

*q in units of output, c in units of total cost.

K is the same as at H, because OK and OH are an identical line and have therefore a common slope.[1]

Suppose now that the specific point H slides along the curve toward the left and takes next the position H' (8, 31.8). Output has been reduced from 9 to 8, and average cost is now

$$\frac{31.8}{8}, \text{ which is } 3.975$$

A new line OH' can be drawn; it will have a smaller slope than OH, and its inclination α' will be smaller than α. Likewise, OH' will cut the curve in a second point K' to the left of H', but K' will be to the right of K.

[1] We can determine the output and cost for K by observing that q and c for K must "satisfy" both the equation for the curve and that for the line OH. These equations are

$$c = 0.2q^2 + 1.5q + 7 \qquad \text{and} \qquad c = \frac{36.7}{9} q$$

(see p. 13, where it appears that the equation of a line of slope m passing through (0, 0) is $c = mq$). Solving these "simultaneous" equations, by substituting c as given by the first equation for c in the second, and then solving the resultant second-degree equation in q, *viz.*,

$$0.2q^2 + \left(1.5 - \frac{36.7}{9}\right) q + 7 = 0$$

yields two values

$$q = {}^{35}\!/_{9} \qquad \text{and} \qquad 9$$

In actual practice, we should probably express the coefficients of the equation in decimals and then solve, getting 3.87 and 9.03 as the two values (approximate) of q. The value 9, of course, belongs to H, and ${}^{35}\!/_{9}$ (approximately 3.9) belongs to K, the corresponding value of c at K being 1284.5/81, which is 15.9 approximately.

In similar manner, we can take a new position H'' (not shown on the chart), still farther to the left than H'; corresponding to this is a line OH'' having a still smaller slope (and therefore representing smaller average cost) and a still smaller inclination α''. Moreover, the corresponding second point K'' would be still farther to the right than K'. Manifestly, as H moves to the left—to successive positions H', H'', etc.—average cost declines, the slope of the line OH becomes smaller, and the "second point" K moves to the right. The chart suggests that presently, in its leftward movement, H will reach a position L where the line barely touches the curve in a single point: H and K come together at L. The line OL is said to be tangent to the curve at L.[1] The inclination β of OL is smaller than the inclination of OH for any of the earlier positions of H (positions to the right of L), and thus average cost is a minimum at L. This geometrical representation of minimum average cost is frequently encountered in texts on economics, but a different treatment is presented below (page 115).

The line OL is called the extreme position of the line OH, the inclination β of OL is the minimum value of the inclination α of OH, and the slope of OL is the minimum value of the slope of OH. Correspondingly, the average cost at L is the minimum value of the average cost at H.[2]

The chart suggests that output q, for L, is approximately 6; careful drawing of the chart on a larger scale might enable us to estimate the q for L to one or more decimal places. But no chart can give an exact determination of q for the position L. For such determination, an analytical method, *i.e.*, a symbolic method, is essential. Such a method requires the use of a more powerful tool of analysis than any discussed above, and development of this tool is the main objective of later sections of this

[1] We have purposely chosen the curve so that tangency will appear in this simple form—of K and H coming together. More complicated forms of tangency can appear, and are discussed in texts on calculus.

[2] The student will now understand why we took the original position of H "far to the right": it was to start out with a position to the right of L. Had we started with H to the left of L, the movement of H into successive positions would necessarily (for this discussion) have been toward the right, and correspondingly K would originally have been to the right of H (and of L) and would have moved to the left. H and K must, for the purpose of this discussion, approach their common position L, from opposite sides of L.

chapter and of the succeeding chapter. Once it is developed, we can return to the problem of the present section and determine the q and c of L exactly (see page 115).

Before proceeding with the main discussion, however, attention can advantageously be given to another view of the situation

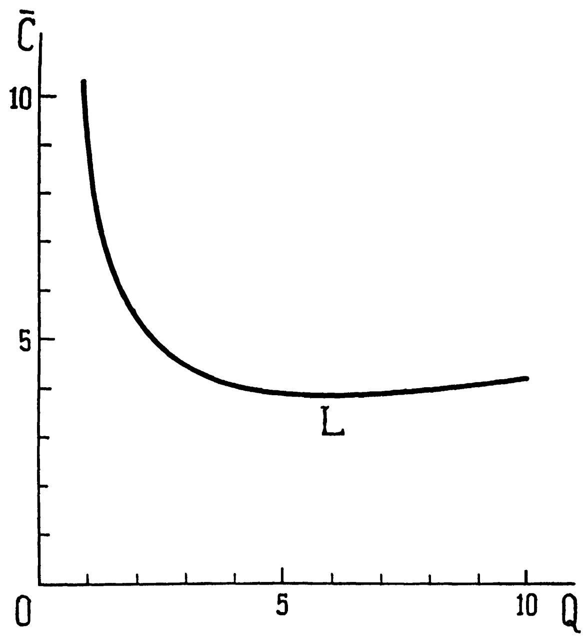

CHART 21.—Average cost, for total-cost function shown in Chart 19.

at L. Corresponding to Equation (16), the average cost equation is

$$\bar{c} = 0.2q + 1.5 + \frac{7}{q} \tag{17}$$

and it is represented by the curve of Chart 21. Chart 21 is the image of Chart 19 (compare a similar case, in Charts 13 and 17 of Chap. II). The curve is inclined downward in its left portion and upward in its right portion. At some intermediate point L the curve reaches its lowest point: there average cost is minimum. The L of Chart 21 corresponds precisely to the L of Chart 20. Here also we can estimate q, for L, from the chart and observe that it is about 6; but again the chart fails to give an *exact* determination of the value of q at L. Once more, exact deter-

mination must await the development of a more powerful tool of analysis.

Several further remarks can helpfully be made about the findings of the present section. In the first place, both the examination of Charts 19 and 20 and that of Chart 21 indicate that, as output starts from zero and increases, (1) average cost

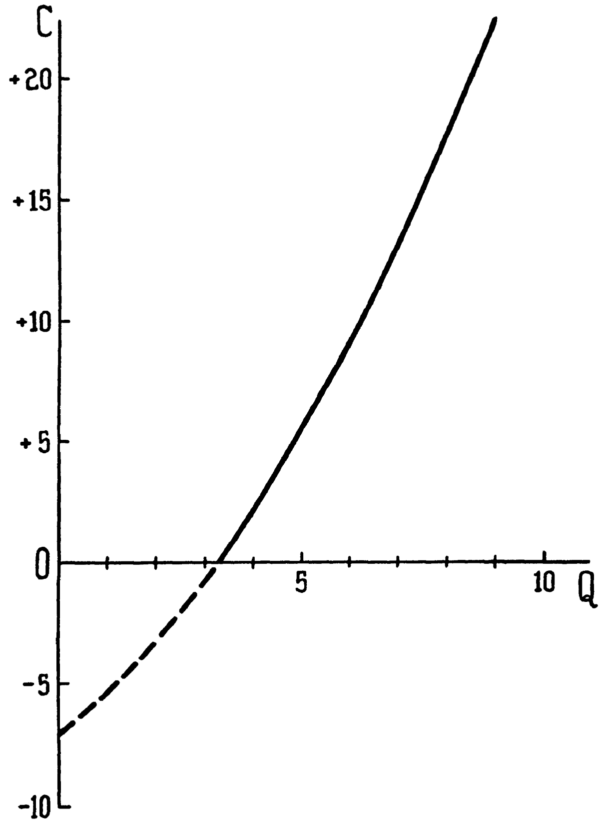

CHART 22.—Total "cost" for function moderately different from that of Chart 19.

at first decreases, (2) presently reaches a minimum, and (3) then increases. We remark now, without explaining in detail, that this is a consequence of the form that has been assumed for Equation (16)—a form that is roughly plausible, for a relatively simple type of costs, from an economic point of view. Had Equation (16) been in somewhat different form, conclusions 1, 2, and 3 might not have been appropriate.

In the second place, an apparently slight change in Equation (16) yields a "cost" curve for no point of which "average cost" is minimum; but this modified form of the cost function admittedly seems not very plausible from an economic point of view. For example, the curve representing.

$$c = 0.2q^2 + 1.5q - 7 \tag{18}$$

appears in Chart 22. Here that portion of the curve for which c is negative, though q is positive, is plotted in dots, for such "costs" do not appear economically plausible; in fact the point where c is zero, and presumably a portion of the curve just to the right of that point, cannot be regarded as economically "realistic."[1] By an examination similar to that applied to Chart 20, we find at once that no point of this curve of Chart 22 corresponds to minimum average cost. In fact, in this case, the line OH never "cuts the curve in a second point" K.[2]

In the third place, consider a borderline case, in which the fixed cost is zero, given by

$$c = 0.2q^2 + 1.5q \tag{19}$$

and shown in Chart 23. Here, manifestly, the line OH always cuts the curve in a second point of intersection, that point being invariably at O no matter where H is located. Hence, as H moves to the left, the line OH does approach an extreme position, and that position is precisely the tangent line at O.

The various charts suggest, but do not establish completely, the following general conclusion about minimum average cost. Any cost situation represented by a function of the type studied

[1] By a somewhat stilted interpretation of the term cost, however, we could conceive of Equation (18) as properly reflecting costs for a particular process or stage of production in a more complex scheme involving other operations; or perhaps such a cost formula might apply to one commodity produced jointly with one or more others. In any case, as stated above, Equation (18) does not appear very realistic as a "cost" formula; it is introduced here merely to bring out the possibility of having the general formula of Equation (12) take a specific form, given by Equation (18), for which there is no minimum "average cost."

[2] A second point of intersection would be found, to the left of O, if we plotted the curve for negative values of q, but we have systematically excluded such negative outputs from consideration. In any case, such second point would move to the left, not to the right, as H moved to the left; H and K would never tend to come together at a single point.

above, with a total cost curve that is concave upward throughout its course [the experimental curves plotted indicate such concavity upward when a of Equation (12) is positive] and which cuts the vertical axis above O [the h of Equation (12) is positive],

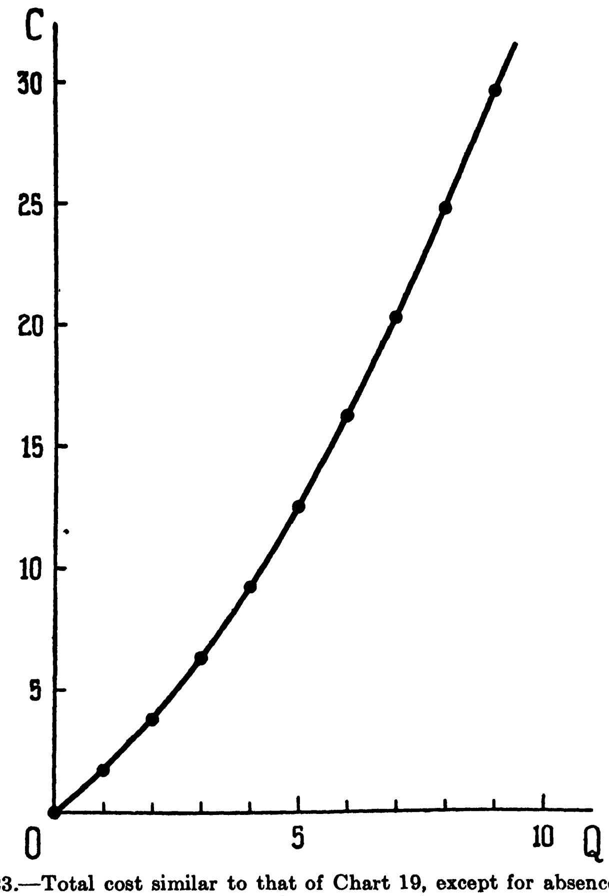

CHART 23.—Total cost similar to that of Chart 19, except for absence of fixed element.

has some particular output, corresponding to a point L, for which average cost is minimum. The exact location of point L can be accomplished by the method developed in Chap. V.

Marginal Cost. We return to the specific formula for total cost

$$c = 0.2q^2 + 1.5q + 7 \qquad (16)$$

and consider it with a different purpose in mind. We are in fact about to consider a carefully defined *change* in the scale of output, starting with the output implied by some point A of the curve

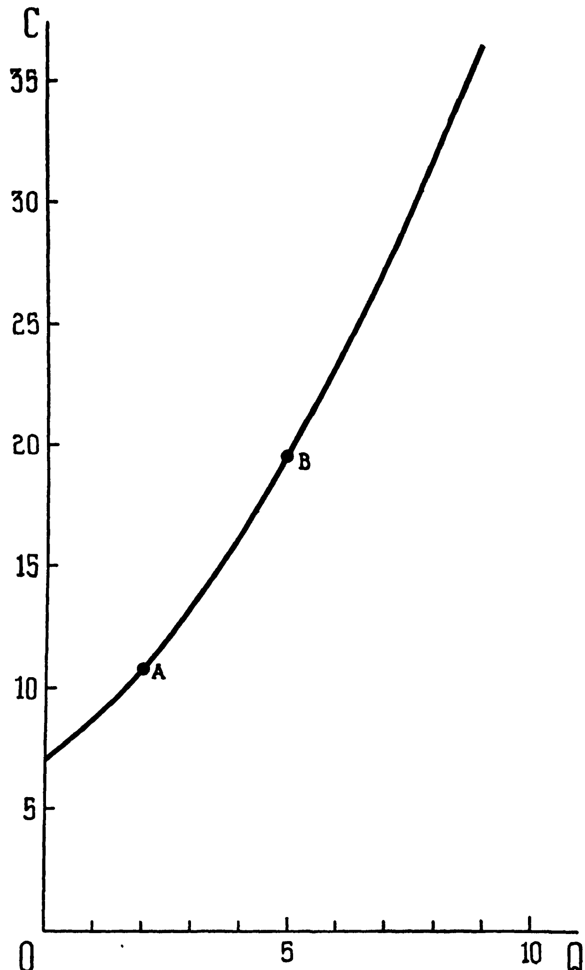

CHART 24.—Repetition of curve of Chart 19.

(reproduced in Chart 24) and then imagining output expanded to that implied by a point B, to the right of A.

Choose A as any point whatever, but preferably not (0, 7), on the curve; to make it numerically specific, let it be (2, 10.8). This point represents the cost situation for output 2. Choose B

as some other point on the curve, to the right of A; for the moment, imagine that B is specifically (5, 19.5). B represents the cost situation for output 5. Suppose the producer started out by operating at output 2, represented by A, and then changed his scale of operations to output 5, represented by B. The change in output is the increase of 3 units, and the corresponding change in cost is the increase of $19.5 - 10.8$, which is 8.7 units. The additional output involved in his change in scale of operations is 3, and the additional cost involved is 8.7. We may therefore say that the average cost of the additional output, the cost per unit of the additional output, is

$$\frac{8.7}{3}, \text{ which is } 2.9$$

This is neither the average cost of the entire output at A, nor the average cost of the entire output at B. The average cost of the entire output at A is

$$\frac{10.8}{2}, \text{ which} = 5.4$$

and at B is[1]

$$\frac{19.5}{5}, \text{ which} = 3.9$$

But the average cost for the 3 additional units of output, secured by changing the scale of operations from A to B, is entirely different from either of these and, in fact, is less than either.[2]

Now suppose that the point B slides along the curve toward A, that the increase in the scale of operations becomes smaller than the three-unit increase examined above. Suppose in fact that

[1] These two results could, of course, have been obtained from the formula for average cost, along the lines discussed in the preceding section. Such formula reads

$$\bar{c} = 0.2q + 1.5 + \frac{7}{q} \tag{17}$$

Substitution of 2 and 5 for q in Equation (17) gives the average cost as

5.4 at A 3.9 at B

[2] If our cost function had differed from Equation (16), and in a very special way, the average cost of the additional output might have been equal to the general average cost of the entire output at A or else at B; but such a highly special case is purposely avoided by our choice of function.

B now falls at (3, 13.3); here the additional output is 1 unit; the additional cost is 13.3 − 10.8, or 2.5 units; and the average cost of the additional output is 2.5 ÷ 1, or 2.5.

Let B move still closer to A, taking on the succession of positions indicated by the values of q in column (1) of Table 7. Column (2) of the table gives the corresponding values of c; columns (3) and (4) give the corresponding values of the additional output [item of column (1) minus 2 units], and the additional cost [item of column (2) minus 10.8]; and column (5) gives the corresponding average cost of the additional output [item of column (4) divided by item of column (3)].

TABLE 7.—COMPUTATION OF THE AVERAGE COST OF THE ADDITIONAL OUTPUT, WHEN SCALE OF OPERATIONS IS RAISED FROM TWO UNITS OF OUTPUT TO q UNITS OF OUTPUT

q (1)	c (2)	Additional output (3)	Additional cost (4)	Average cost of additional output (5)
5	19.5	3	8.7	2.9
3	13.3	1	2.5	2.5
2.5	12	0.5	1.2	2.4
2.1	11.032	0.1	0.232	2.32
2.01	10.82302	0.01	0.02302	2.302

The final column suggests, but does not prove, that the average cost of the additional output becomes smaller and smaller as B gets closer and closer to A; it also suggests that the average cost of the additional output is approaching a definite value, which appears to be about 2.3. When a function y approaches a definite value y_1, as the independent variable x approaches a value x_1 as its limit, y_1 is called the *limit* of y (see pages 41 and 42).[1]

This limit of the average cost of the additional output, as the additional output approaches zero, is defined as the *marginal cost.* Strictly, the term might better have read "marginal cost per unit," but the term marginal cost is customarily accepted in this sense. We may well pause to observe that this concept

[1] This definition can be more strictly stated: If y is a function of x and if the difference between y and a constant y_1 can be made smaller than any previously assigned amount however small by taking x closer and closer to x_1, y_1 is called the limit of y as x approaches x_1.

appears on its face to be highly artificial: the average cost of the additional output when the additional output is nil has of course no reality, but this is not implied in the limit concept. The *limiting value* of the average cost of the additional output, as the additional output approaches zero, is a very real number, and appears in the case under study to be approximately 2.3. The concept of marginal cost, which is defined in the limit termi-

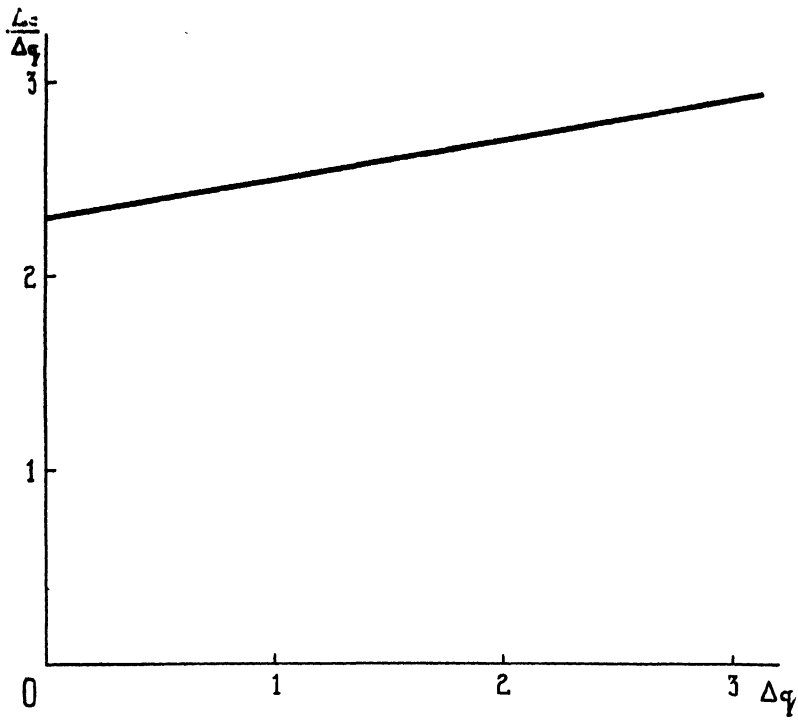

CHART 25.—Average cost of an additional output beyond *A* of Chart 24.

nology, has in truth basic significance in the treatment of cost problems in economic theory; a corresponding concept of a marginal magnitude is of similar importance in many other branches of economic theory. The producer may or may not be aware of his marginal cost; but, assuming that certain competitive conditions prevail, in his adjustment of output to market conditions he acts, perhaps unconsciously, in accord with the principle that marginal cost tends to equal price, that output tends to expand until marginal cost equals price. Economic theory

develops this principle as a fundamental tenet of the doctrine of value; the analysis of the concept is for that reason, as well as on account of its mathematical implications, an essential step in our examination of the *cost* function.

The foregoing method—of choosing specific locations for the second point B, and calculating the corresponding numerical results for the average cost of the additional output—does not, however, yield a determinate result for the limiting value of the average cost of the additional output, as B actually falls upon A; for then both additional output and additional cost become zero, and the ratio of zero to zero has no meaning.

We might make a supplementary chart, or image, in which the horizontal axis represents additional output, labeled Δq, and the vertical axis represents average cost, labeled $\Delta c/\Delta q$, of the additional output, such as Chart 25. Here the direction of the curve at the left end suggests 2.3 as the limiting value when additional output approaches zero. But the chart does not unmistakably show 2.3 as the limiting value; one observer might estimate it as 2.31 and another as 2.295, etc. The chart gives no precise determination; as pointed out above, no point can be calculated and plotted for the case when additional output is zero.

Exact Determination of Marginal Cost. A more powerful method is therefore needed, if a precise determination of the limiting value is to be secured. To develop this more powerful method, we abandon the specific numerical designation of points A and B and represent them by (q_1, c_1) and (q_2, c_2), respectively (Chart 26 in which the curve of Chart 24 is reproduced).

The additional output, which we shall now represent by the symbol Δq, is

$$\Delta q = q_2 - q_1 \tag{20}$$

and the additional cost, represented by Δc, is[1]

$$\Delta c = c_2 - c_1 \tag{21}$$

Another form of Equation (20) is

$$q_2 = q_1 + \Delta q \tag{20a}$$

[1] The symbol Δq (and, likewise, Δc) is a compound symbol (see note 1, p. 13)—it is not Δ "times" q any more than $\sqrt{2}$ is $\sqrt{\ }$ "times" 2. It is customarily read "increment in q." The student should avoid calling it a "small change in q"; it need not be small.

Moreover, as A and B are points of the curve, (q_1, c_1) and (q_2, c_2) must satisfy Equation (16)

$$c_2 = 0.2q_2^2 + 1.5q_2 + 7 \tag{16a}$$
$$c_1 = 0.2q_1^2 + 1.5q_1 + 7 \tag{16b}$$

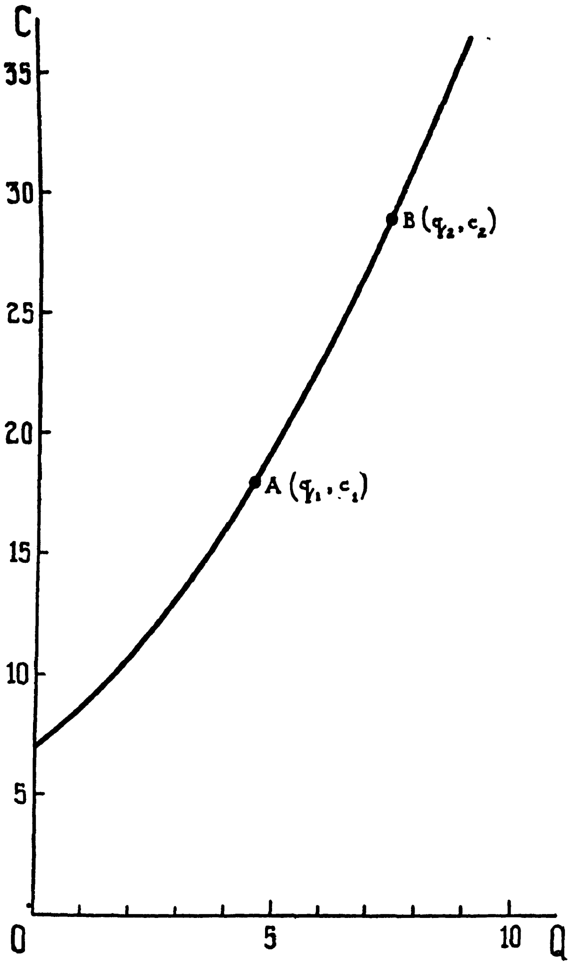

CHART 26.—Repetition of curve of Chart 19.

and the first of these becomes, by using Equation (20a),

$$c_2 = 0.2q_1^2 + 0.4q_1\Delta q + 0.2(\Delta q)^2 + 1.5q_1 + 1.5\Delta q + 7 \tag{16c}$$

Subtracting Equation (16b) from Equation (16c) yields

$$\Delta c = c_2 - c_1 = 0.4q_1\Delta q + 0.2(\Delta q)^2 + 1.5\Delta q \tag{21a}$$

The average cost of the additional output Δq is the ratio of the additional cost Δc to Δq, and, by Equation (21a), it is

$$\frac{\Delta c}{\Delta q} = 0.4q_1 + 0.2\,\Delta q + 1.5 \tag{22}$$

From Equation (22) we can calculate, by substitution for q_1 and Δq, the average cost of the additional output for any assumed values of the original output q_1 and the additional output Δq. Thus, taking q_1 as 2 and Δq as 3 to correspond to our first numerical case above, we get

$$\begin{aligned}\frac{\Delta c}{\Delta q} &= 0.4 \times 2 + 0.2 \times 3 + 1.5 \\ &= 2.9\end{aligned}$$

as before. In similar manner, we could reproduce all the numerical results found above. Equation (22) is a perfectly general formula for getting the average cost of the additional output—for the cost situation of Equation (16).

Moreover, Equation (22) enables us to take a step not possible by the earlier numerical method: we can now determine the exact limiting value of $\Delta c/\Delta q$ as Δq approaches zero. Mathematicians have a standard symbol for the limiting value, or limit, reached under specified circumstances; this symbol is merely a kind of shorthand expression for a verbal phrase, just as $\sqrt{\ }$ in $\sqrt{2}$ means "the square root of" 2. The symbol is

$$\lim_{\Delta q = 0} \frac{\Delta c}{\Delta q}$$

which is read "the limit of $\Delta c/\Delta q$, as Δq approaches zero." The notation $\Delta q = 0$ below the main symbol is thus the "phrase" describing the manner in which the limit of the ratio is approached; it is approached in this problem by letting Δq approach zero.

From Equation (22) we find

$$\lim_{\Delta q = 0} \frac{\Delta c}{\Delta q} = 0.4q_1 + 1.5 \tag{23}$$

because the term $0.2\,\Delta q$ approaches zero as Δq approaches zero. Equation (23) gives, by definition of marginal cost, the marginal

cost for output q_1. If we choose for marginal cost the symbol c_m, we have

$$c_m = 0.4q_1 + 1.5 \tag{24}$$

For the particular scale of operations represented by the numerically specific point A, for which output is 2, Equation (24) gives for the marginal cost

$$\begin{aligned} c_m &= 0.4 \times 2 + 1.5 \\ &= 2.3 \end{aligned}$$

as we estimated from Table 7. The result, however, is now no longer an estimate, it is exact.

The process of evaluating the limit of $\Delta c/\Delta q$, in symbolic terms, thus affords a more powerful method than the cut-and-try process of numerical substitution first attempted. The limit concept is not only real, but the exact value of the limit can be determined. The introduction of this limit concept as a tool of analysis is the most important single step we shall take in elaborating our elementary mathematical equipment for the study of economic problems. Contained within it is the root idea of the differential calculus, and much of the supposed mystery of calculus will disappear if the student will get clearly in mind the process, as single steps and as a systematic method, leading up to Equation (24).

So important is this method, with the reasoning upon which it rests, that some further illustrations of the limit concept as applicable in economics, and of the process of evaluating the limit, will now be given. Although this direct process of evaluating the limit will be replaced in the next chapter by a more elegant, and in some respects simpler, process, the method here used so closely conforms to the basic reasoning upon which the more elegant process rests that the reader will find the illustrations here presented immensely helpful.

Demand Curves, Individual. For the purpose of presenting these further illustrations, we turn from cost theory, which has thus far provided all our illustrations, to another field of economic theory. Consider an individual consumer, having a stated income, with particular reference to his demand for a specific commodity at a stated time. Facts observed from common experience are (1) that the quantity q of any particular com-

they can nevertheless sometimes be treated as if they were quantities. Of this, too, we have already had an example (see page 151). In this book (see especially page 88) the derivative has been defined as an operational symbol. dy/dx is not a fraction, but d/dx indicates an operation that is to be carried out upon y. The student has nevertheless been taught to use the differential *notation*, which amounts to treating this operational symbol as if it were a fraction with dy and dx as numerator and denominator. What we are about to do now is nothing but another step in the same direction; just as the differential notation led us to treating dy and dx as quantities that may be dealt with according to the rules of ordinary algebra, so we are now going to treat the symbol d/dx itself as if it were such an algebraic magnitude.

For the sake of convenience, we introduce the notations

$$\frac{d}{dx} \equiv D_x, \qquad \frac{d^2}{dx^2} \equiv D_x^2, \qquad \frac{d^3}{dx^3} \equiv D_x^3 \cdots$$

and throw Equation (83) into the form

$$(a_0D_x^n + a_1D_x^{n-1} + a_2D_x^{n-2} + \cdots + a_n)y = 0 \tag{84}$$

Now we take the decisive step. For the moment, we treat the $D_x^n, D_x^{n-1}, D_x^{n-2}, \ldots$ exactly as if they were unknown quantities in an algebraic equation of degree n; and we treat the numbers $n, n - 1, n - 2, \ldots$ as if they were the exponents of those quantities and not, what they actually are, indicators of the order of the differentiations to be performed. That is to say, we treat, for instance, D_x^2 as if this symbol meant some quantity m that is to be raised to the second power instead of meaning that some variable, say y, is to be differentiated twice. In order to emphasize the nature of the procedure and to avoid misleading implications, we shall formally set up an *auxiliary equation* in m which is, momentarily, to replace the bracketed expression in Equation (84). We write

$$a_0m^n + a_1m^{n-1} + a_2m^{n-2} + \cdots + a_n = 0 \tag{85}$$

We then solve this equation, getting a number of values for m that is equal to the order of the equation [the order of the operational equation (84), the degree of the algebraic substitute (85)],

concerned. We shall therefore present our charts with q measured horizontally and p measured vertically and speak of p as "a function of q," though it is equally true that q is a function of p.

The functional relation for a commodity A is sometimes stated in tabular form, as in Table 8. The corresponding points are plotted in Chart 27; they are not joined by a curve or by line segments, although presumably the individual's demand (waiving the fact that ultimately the physical units of a particular com-

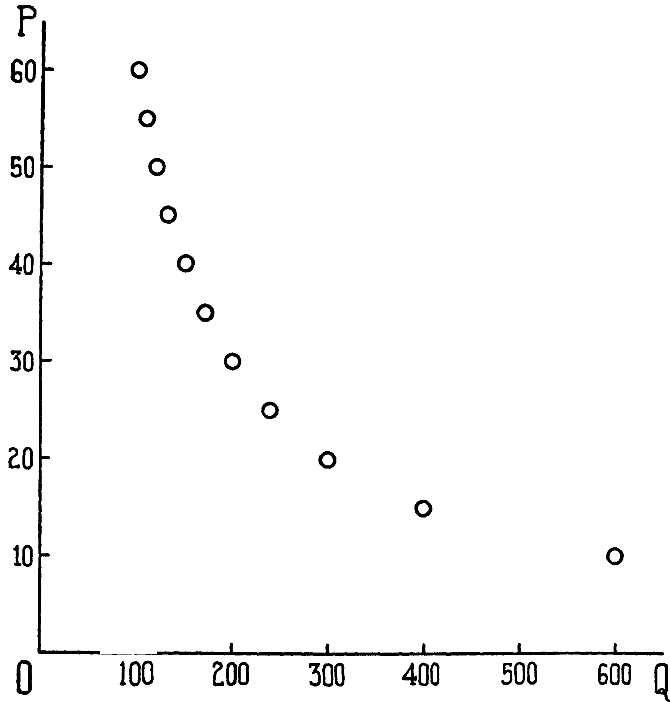

CHART 27.—Demand schedule, selected points only, given by Table 8.

modity, in case of most commodities, become indivisibly small, see note 1, page 21) is in fact a continuous function of price, and correspondingly price is a continuous function of quantity.

The functional relation can also be stated in symbolic form, and here we make a succession of assumptions. As in the case of our earlier cost functions, the exact basis on which any one of these assumed formulas rests need not be known, and we need not even specify whether the basis is a priori or empirical. That the assumed functions broadly satisfy the qualitative

description covered by the three points above (page 58) is sufficient for the present purpose.

The first assumption is that the relation between price and quantity is a straight line, inclined downward-to-the-right. Such

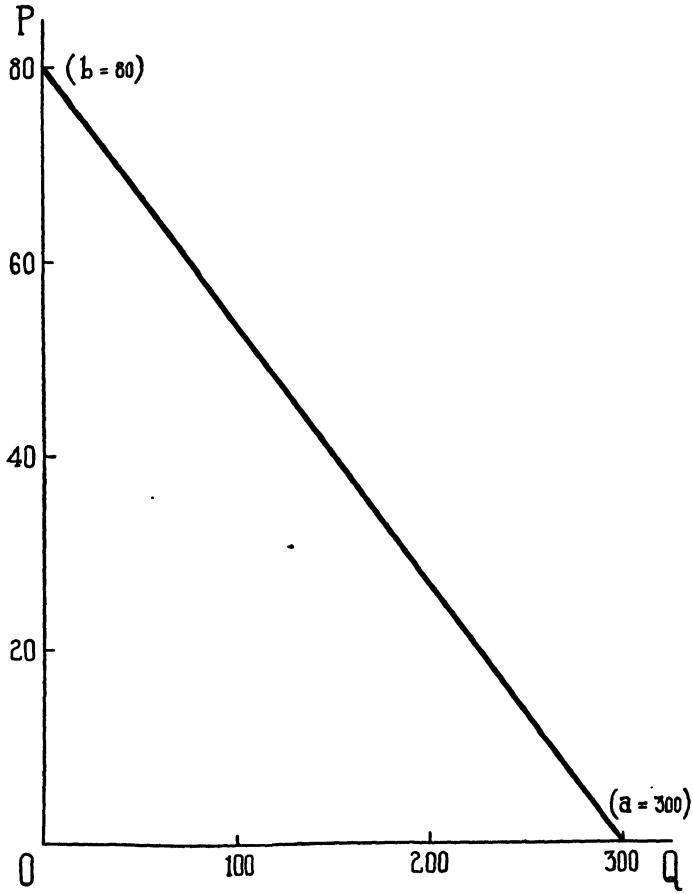

CHART 28.—Assumed linear-demand line.

a line may be taken as having intercepts a and b on the OQ and OP axes, as in Chart 28. The equation, in implicit form, of that line is[1]

$$\frac{q}{a} + \frac{p}{b} = 1 \tag{25}$$

[1] The reader will find an earlier discussion of the straight-line equation in terms of its intercepts in footnote 2, p. 24. He can "verify" the present

Actually, the equation is represented by an endless line extending left of OP and above OQ' (the leftward extension of OQ), and also below OQ and right of OP', but we use only the portion between OQ and OP, as alone having economic significance. Manifestly, this functional relation accords broadly with the three-point qualitative requirements in the text above: quantity

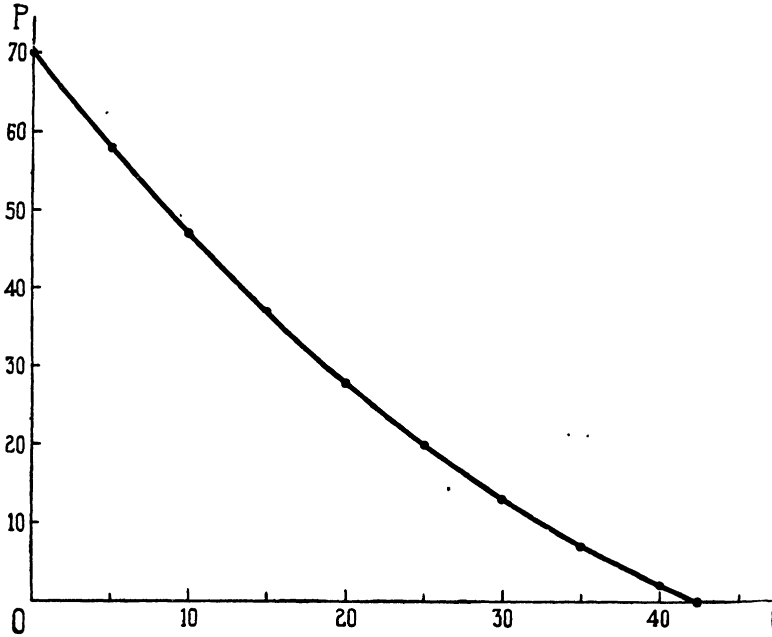

CHART 29.—Assumed demand curve of parabolic type.

depends on price, quantity increases as price decreases, quantity is not indefinitely large for zero price and price is not indefinitely large for zero quantity.

A second assumption is that the functional relation is represented by a limited portion, the left portion that falls between the OQ and OP axes, of the curve belonging to

$$p = a + bq + hq^2 \tag{26}$$

---

equation by observing that when p is zero q is a (300 in Chart 28), and when q is zero p is b (80 in Chart 28).

In the actual plotting of Chart 28, the general formula, Equation (25), is made specific by assuming the numerical values 300 and 80 for a and b, respectively.

(Chart 29).[1] This curve, called a *parabola*, rests upon a second-degree function similar to that examined in cost theory, Equation (16). Again, the curve of Chart 29 meets the three qualitative requirements stated in the text above. A distinguishing feature of the present curve is that it is concave upward; while price decreases as we pass to the right, as quantity increases, it (price) decreases at a diminishing rate. This notion of rate of decrease (or increase) has already been encountered in the study of cost functions and will receive careful attention below (page 97).

TABLE 9.—PAIRS OF VALUES OF q AND p FOR THE EQUATION $p = 70 - 2.5q + 0.02q^2$

Quantity q	Price p	Quantity q	Price p
0	70	25	20
5	58	30	13
10	47	35	7
15	37	40	2
20	28	42.35*	0

* Approximate.

An alternative form of the second assumption, resting on the same general formula given in Equation (26) but having different values of the constants and in particular a negative b and negative h, is illustrated in Chart 30.[2] This curve again meets the

[1] Equation (26) is rendered specific, for purposes of plotting, by taking the numerical values 70, -2.5, and 0.02 for the constants a, b, and h, respectively. Plotted pairs of values are shown in Table 9. To yield such a figure as Chart 29, b must be negative and h positive.

Actually, Equation (26) is represented by an endless curve that extends upward-to-the-left of OP, drops somewhat below OQ (at the right of the plotted portion), and then rises above OQ and sweeps off upward-to-the-right. Economic considerations obviously require that we confine the curve to the downward-to-the-right portion (the left portion) between OQ and OP, as shown in the chart.

[2] For this chart we assume the numerical values 70, -0.5, and -0.005 for a, b, and h, respectively; the plotted pairs of values appear in Table 10.

Actually, Equation (26) is represented, for negative b and negative h, by an endless curve which, in passing to the left of OP, rises for a moderate distance and then turns downward and sweeps off indefinitely toward the lower left, and which continues below OQ at the right end and sweeps downward endlessly to the right. Economic reality requires us to confine the curve, as plotted, between positive OQ and positive OP.

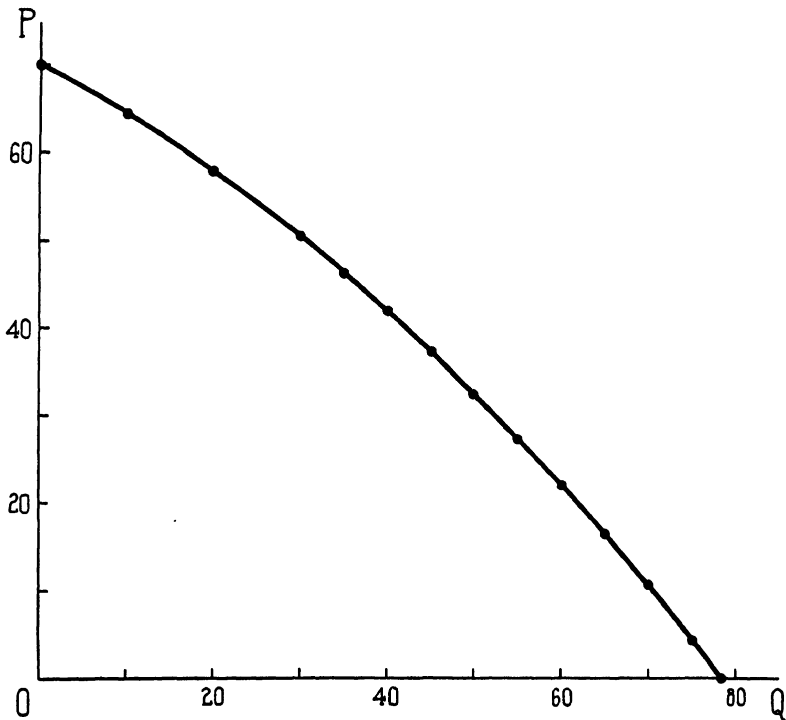

CHART 30.—Demand curve of parabolic type, second example.

three qualitative requirements, but it is concave downward. As we pass to the right, price decreases, but it decreases at an increasing rate. Although such a curve is not necessarily

TABLE 10.—PAIRS OF VALUES OF q AND p FOR THE EQUATION
$p = 70 - 0.5q - 0.005q^2$

Quantity q	Price p	Quantity q	Price p
0	70	50	32.5
10	64.5	55	27.375
20	58	60	22
30	50.5	65	16.375
35	46.375	70	10.5
40	42	75	4.375
45	37.375	78.45*	0

* Approximate.

unrealistic, it is less likely to arise in true demand situations than a curve shaped like that of Chart 29.

A third assumption worthy of study is that the functional relation is

$$pq = k, \qquad \text{with } k \text{ constant}$$

This is in implicit form, but can be solved explicitly for p to yield

$$p = \frac{k}{q} \tag{27}$$

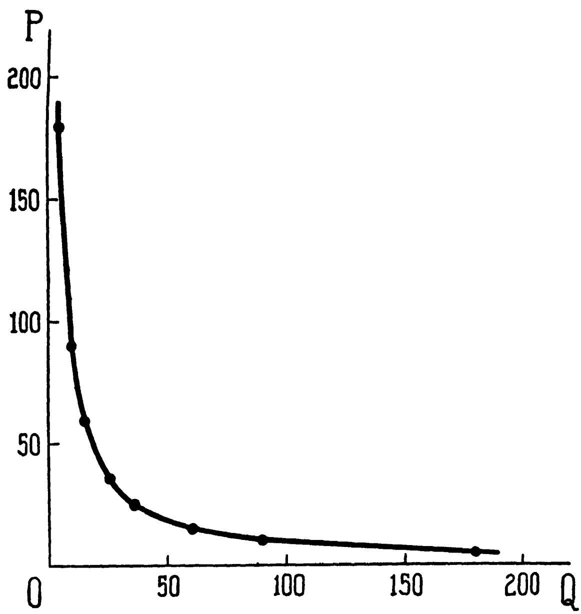

CHART 31.—Demand curve of hyperbolic type: curve of unity elasticity.

Table 11 furnishes pairs of values for the specific case $k = 900$, and the corresponding curve is presented in Chart 31. Although

TABLE 11.—PAIRS OF VALUES OF q AND p FOR THE EQUATION $pq = 900$

Quantity q	Price p	Quantity q	Price p
5	180	36	25
10	90	60	15
15	60	90	10
25	36	180	5

the curve is plotted only for q between 5 and 180, it actually extends indefinitely upward as q approaches zero and indefinitely to the right as p approaches zero. The entire curve, throughout any range we desire to calculate, lies between OQ and OP. At no point does the curve cut or touch OQ or OP; this may be seen by substituting q (or p) as zero in Equation (27) and attempting to calculate the corresponding p (or q). The "result" is k "divided by" zero (900 "divided by" zero), and we formed the habit in elementary algebra of calling such a result *infinite*.[1] A careful statement of the situation reads: By taking q (or p) sufficiently small (but never actually equal to zero), we can make p (or q) larger than any number however large which we care to assign in advance. Thus, if we want to make p larger than 9 million, we have merely to take q smaller than 1/10,000.

The curve of Chart 31 manifestly meets the first two of the qualitative requirements listed above (page 58); but, as it does not admit of q or p ever becoming zero, it does not meet the third requirement. It is therefore not a curve likely to reflect an individual's demand for an ordinary commodity in any realistic sense; but it has such theoretical importance, and is so close to certain realistic cases, that its study is advantageous.

[1] More loosely, we may have become accustomed to call it "infinity." This word is objectionable (see p. 20) and should constantly be shunned by the student, for its use involves the danger of coming to regard an infinite number as having some definite size, just as 7 or 816 has a definite size. This notion is essentially wrong: an infinite number has no definite magnitude, its basic quality is that it is indefinitely large when compared with ordinary numbers. Thus, strictly, we cannot say "900 divided by zero equals infinity," for actual division by zero is ruled out. We must say that the ratio of 900 to x, as x becomes indefinitely close to zero, becomes indefinitely large, becomes infinite. To be very careful, we must say that, by taking x smaller than some specified small number (very small fraction, presumably), we can render the ratio larger than any previously assigned large number, however large. The student may regard all this as a mere quibble, but it is said to help him get out of his mind the fallacious notions that (1) actual division by zero is possible, and (2) an infinite number has a definite size. In analyses involving limits we are often concerned with rendering certain variables exceedingly small, with allowing them to "approach" zero, and with rendering certain other variables exceedingly large, with permitting them to become infinite, and we must never forget that in these cases the variable does not actually take on the value of the limit that it approaches.

For the present, we note that it is concave upward, price decreases as quantity increases, but always at a diminishing rate. Moreover, the curve has a peculiar sort of symmetry; if we imagine the chart folded along a line bisecting the angle POQ, the two halves of the curve fall into exact coincidence with each other.[1]

Marginal Utility. The older theory of demand rests upon the hypothesis that the total taking by an individual, of a particular commodity at a stated time, contributes in a measurable degree to the total utility enjoyed by him. The utility of a commodity to an individual represents the intensity of his desire for it and reflects its capacity to afford him certain satisfactions—in a psychological, ethical, or other sense. Although "satisfaction" does not readily lend itself to measurement, economic theory used to assume that the related phenomenon utility, or at least a change in utility, is measurable and that its measurement is in pecuniary terms. Just how the utility of a commodity to an individual is related to the satisfactions derivable by him from it does not admit of precise determination. The relation can, however, be satisfactorily covered for theoretical purposes by assuming, as seems plausible, that there is for the individual a utility function u (u being a function of the quantity of the commodity) having the following properties: (1) it remains unchanged if the satisfactions yielded by the commodity remain unchanged, (2) it varies in the same direction (increases or decreases) as the satisfactions vary (increase or decrease), and (3) the larger of two variations in u is associated with the larger of two variations in satisfactions. These properties manifestly establish only certain qualitative relations between u and satisfactions.[2] We take it for granted, however, that the

[1] The student familiar with analytic geometry will recognize this as a property of the *equilateral hyperbola*, of which Equation (27) is the formula. He will know also that there is another and similar branch of the curve, lying between the negative axis OQ' and the negative axis OP'; but this branch has been ignored here as obviously without economic significance.

[2] For more elaborate statement, see A. L. BOWLEY, *Mathematical Groundwork of Economics*, p. 1, Oxford University Press, New York, 1924. See also GARVER and HANSEN, *op. cit.*, pp. 134–151. Many students will, however, be aware of the fact that the approach chosen in the text herewith no longer commands the approval of the majority of economists and has been replaced by another approach that works with "indifference maps." Still later developments have produced a method that also does away with the latter. But we have here chosen the utility approach because it affords the

individual does appraise his satisfactions with tolerable accuracy in terms of utility; we assume that his measurements of utility consist in making his appraisals of desiredness and in reflecting them in his economic behavior, particularly as respects his demand for the commodity. He makes decisions, whether to acquire or forego an additional unit (or small quantity) of a

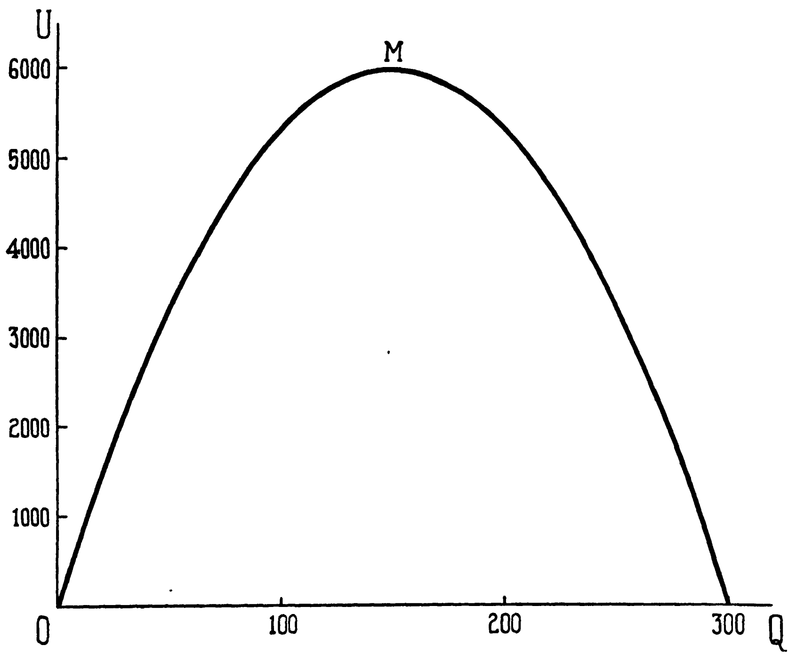

CHART 32.—Total-utility curve, parabolic type.

commodity when he already has a supply of the commodity, and these decisions result in actions that reflect measurements of his utility in pecuniary terms.

For present purposes, we shall assume that the utility u can be expressed as a function of the quantity q, of the commodity, that the individual actually takes (see page 150 for discussion of a tentative "derivation" of a utility function and of certain difficulties incident thereto). Thus, we assume

$$u = f(q)$$

---

best opportunity for displaying our mathematical concepts in the simplest way.

and, to supply a particular illustration, we further assume that, for a particular individual and a particular commodity in the circumstances existing at a particular time, the function has the general form

$$u = bq - \frac{b}{a}q^2 \tag{28}$$

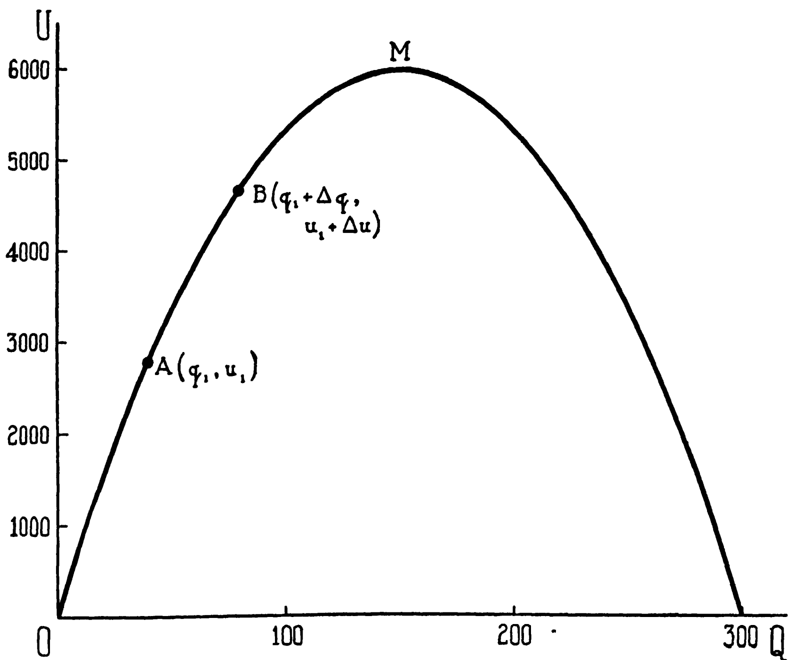

CHART 33.—Change in utility for curve of Chart 32.

where b and a are positive constants. To render this numerically specific, choose b and a so that

$$u = 80q - \tfrac{4}{15}q^2 \tag{29}$$

and the function can then be plotted (by use of a table of values, not reproduced) as in Chart 32.[1]

Suppose now that we consider the change in utility associated with a change in quantity. Let point A, Chart 33 (which exactly

[1] Actually, the curve extends downward-to-the-left of OU and also downward-to-the-right below the right end of OQ. Only the positive portion of the curve is shown, and only the left (rising) half of that portion has economic meaning.

reproduces the curve of Chart 32), represent the relation between u and q for any chosen q (called q_1) whatever, except that A should not be chosen at O or at the right of the highest point M.

Suppose now that the individual's taking increases from q_1 to $q_1 + \Delta q$, and that the new situation is represented by point B $(q_1 + \Delta q,\ u_1 + \Delta u)$. The corresponding increase Δu, in u, may be regarded as the increase in utility associated with the increase Δq, in quantity. The corresponding *average* increase in utility, per unit of increase in quantity, is

$$\frac{\Delta u}{\Delta q}$$

and, as [using the general equation (28) rather than the specific equation (29) chosen for plotting]

$$\begin{aligned}
\Delta u &= (u_1 + \Delta u) - u_1 \\
&= b(q_1 + \Delta q) - \frac{b}{a}(q_1 + \Delta q)^2 - \left(bq_1 - \frac{b}{a}q_1^2\right) \\
&= b\,\Delta q - 2\frac{b}{a}q_1\,\Delta q - \frac{b}{a}(\Delta q)^2, \\
\frac{\Delta u}{\Delta q} &= b - 2\frac{b}{a}q_1 - \frac{b}{a}\Delta q.
\end{aligned}$$

If B is now allowed to slide along the curve toward A, Δq will approach zero and

$$\lim_{\Delta q = 0} \frac{\Delta u}{\Delta q} = b - 2\frac{b}{a}q_1 = u_m. \tag{30}$$

This limiting value u_m is called the *marginal utility* at point A (or for quantity q_1).

Thus, marginal utility at a point A of the utility curve is defined as the limiting value of the ratio of the increase in utility, associated with an increase in quantity, to that increase in quantity, as the increase in quantity approaches zero as a limit Neglecting the subscript 1 of q, we may regard Equation (30) as stating u_m as a function of q; it is plotted (for the selected a and b) in Chart 34. Once more the image relationship appears between two charts: Chart 34 is the image of Chart 33, the relation between them being specified by the definition of marginal utility.

We observe,[1] from Equation (30), that u_m is positive for q_1 less than $a/2$ and is negative for q_1 greater than $a/2$. In fact, Chart

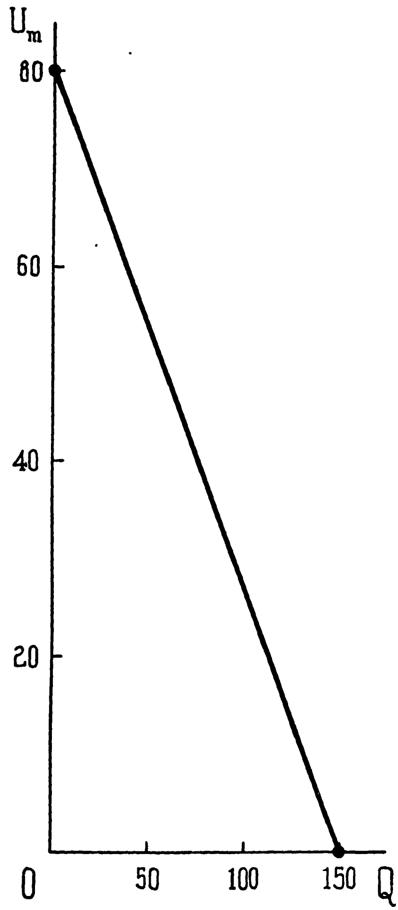

CHART 34.—Marginal utility for total-utility function of Chart 32.

32 showed that, for some point M not precisely known in terms of q, u is maximum; to left of M, u increases with q, to right of M, u decreases as q increases. We shall find (page 118) that

[1] The foregoing analysis, though worked out for points on the curve left of M, is mathematically valid for the right portion also. Economic considerations imply that negative values of u_m are not realistic; hence the portion of the curve of Chart 32 to right of M is not economically significant; hence M is not strictly a maximum, for utility, in the sense defined below, p. 150, except perhaps in a Crusoe economy where a good may conceivably be present in so great a quantity that part of it constitutes a nuisance.

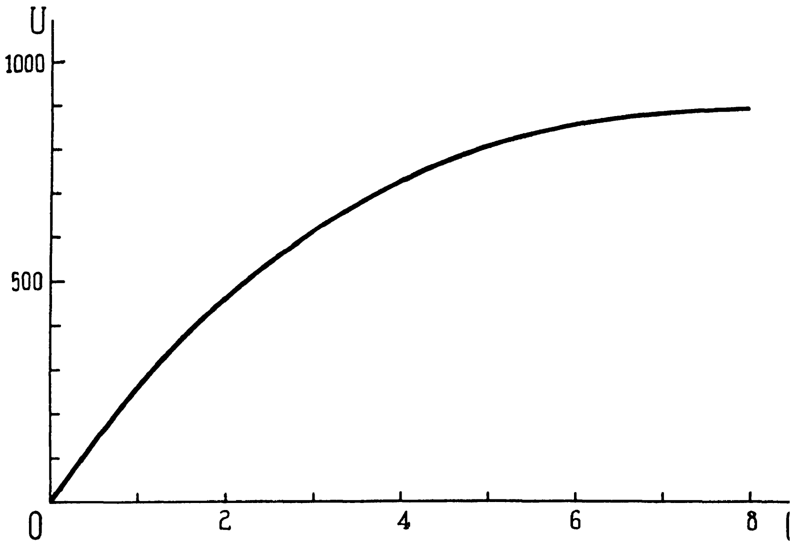

CHART 35.—Total-utility curve, cubic-parabola type.

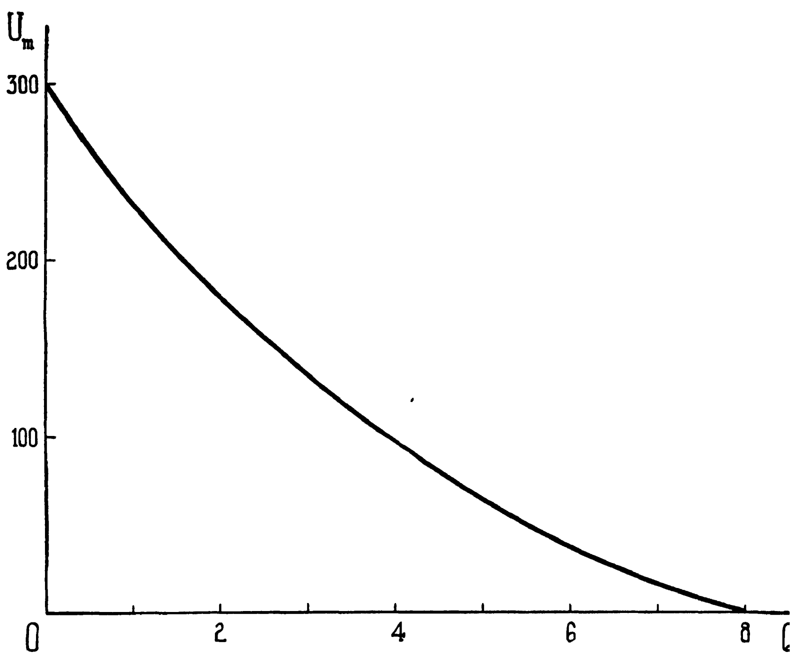

CHART 36.—Marginal utility for total-utility function of Chart 35.

this maximum point M, for u, occurs for precisely the value of q, *viz.*, $a/2$, which makes u_m zero. For the present, we merely assert: left of M, marginal utility is positive, right of M, it is negative.

Consider a more complicated utility function

$$u = aq + bq^2 + hq^3 \tag{31}$$

where a, b, and h are constants. Here, following a like procedure

$$u + \Delta u = a(q + \Delta q) + b(q + \Delta q)^2 + h(q + \Delta q)^3$$

$$\Delta u = a\,\Delta q + 2bq\,\Delta q + b(\Delta q)^2 + 3hq^2\,\Delta q + 3hq(\Delta q)^2 + h(\Delta q)^3$$

$$\frac{\Delta u}{\Delta q} = a + 2bq + b\,\Delta q + 3hq^2 + 3hq\,\Delta q + h(\Delta q)^2$$

$$u_m = \lim_{\Delta q = 0} \frac{\Delta u}{\Delta q} = a + 2bq + 3hq^2 \tag{32}$$

where the subscript 1 has been omitted from q throughout.

Charts 35 and 36 show the utility function, equation (31), and corresponding marginal utility function, equation (32), for a, b, and h chosen as 288, -30, and 1. (The tables of values are not reproduced, but can be constructed by the reader.) From these charts it will be seen that this utility function fulfills the conditions ordinarily imposed upon such functions, only to the point where $q = 8$.

A third assumption as to the form of the utility function, having some importance for certain problems of economic theory, takes

$$u = k \log q \tag{33}$$

where k is constant.[1] The marginal utility in this case cannot

[1] The right member of this equation is to be read "k times log q"; log q is a compound symbol, meaning "the logarithm of q." For the benefit of readers not familiar with logarithms, we note that "the logarithm of a number x to the base b is the exponent of that power of b which equals x." Thus, if

$$y = \log_b x$$

(ordinarily the base of a logarithm is indicated, as here, by attaching it as a subscript to "log") the above definition means

$$b^y = x$$

These two equations—the first, which expressed y explicitly in terms of log x, and the second, which expresses x explicitly as equal to b to the y

readily be calculated by the direct application of the limit process along the foregoing lines, and discussion of this interesting case would be facilitated by use of the derivative concept (page 87).

Demand Curves, Collective. The demand curves discussed above pertain to an individual consumer. The present section deals with a generalized demand curve, representing the collective demand, for a particular commodity at a stated time, resulting from the compounding of the demands of all individuals whose demand is felt in the market to which the collective demand curve (or function) pertains.[1] Such a collective demand function,

---

power—are thus alternative forms of the same functional relation.

When the symbol "log" is used without any subscript, the base is understood as having one of two standard values, either 10, when logarithms are being used as a tool for ordinary computations, or the fixed constant e (called the Napierian base, and equal to 2.71828183, disregarding further decimals), when logarithms arise in ordinary theoretical problems. Which of these two bases is to be understood in a given case will ordinarily be obvious; otherwise the base, 10 or e as the case may be, would be indicated by the subscript. In all cases where the base is neither 10 nor e, it should be indicated by a subscript.

Understanding that the base is e in Equation (33), the reader will note that that equation can be solved explicitly for q

$$q = e^{u/k}$$

When logarithms are used as a tool in ordinary computations, the base 10 is usually adopted. Elaborate tables, tables of logarithms, have been compiled, giving (to the desired number of decimal places) the logarithms to the base 10 of successive numbers. By use of such a table, various arithmetical operations—such as multiplication, division, raising to powers, extracting roots—can be carried out much faster and more easily than by long-hand arithmetic. The results of such computations by logarithms are generally not exact, but any desired degree of approach to accuracy can be obtained by using a table of logarithms that is carried to enough decimals. For the principles underlying computation by logarithms, and exposition of the methods actually used, the student should consult a text on college algebra.

We remark finally that the ordinary slide rule is merely a device for making certain computations (approximate) by a mechanical use of the principles of logarithms.

[1] We refrain from attempting a precise definition of the term *market*, but the reader will find pertinent discussions in economic treatises. Moreover, we do not specify how the demands of individuals are "compounded" in the market. An obviously simple assumption is that they (the quantities demanded) are merely added, for each price; but the reader will suspect that

when stated in explicit terms, gives p equal to an expression involving q, where q is now the aggregate quantity taken by all individuals, in the stated market, at price p.

The qualitative characteristics of such a collective demand function are largely similar to those stated above (page 58) for an individual demand function. But the force of the third characteristic—that quantity becomes zero for a moderately high price, and that quantity is only moderately large for zero price—is here somewhat reduced. In other words, there is a greater likelihood that the collective demand curve will not cut the OP axis, if at all, until it rises far above O, or the OQ axis, if at all, until it passes far to the right of O. This modified property of the collective demand function results from various considerations, which the reader will find elaborated in economic treatises; but a principal reason is that inequality of income among individuals renders the individual curves, to be compounded into the collective curve, very different one from another.

Moreover, the reader must bear in mind that the very shapes of the individual demand curves, going to make up the collective curve, may well differ one from another. Likewise the shape of the resultant collective demand curve need not be similar to all, or even any, of the shapes of the constituent individual curves. Thus, for a given commodity in a stated market, some individuals may have demand curves of the simple straight-line type [Equation (25)] with various values of its constants, others of the parabolic type [Equation (26)], others perhaps of the hyperbolic type [Equation (27)], and still others of types not discussed above.[1] Correspondingly, the resultant collective demand curve would have a mathematical form different from some of these, and might differ from all of them. Again we remark that, for present purposes, the precise manner in which a collective demand function arises from its individual constituents needs not be known; we merely assume that the collective demand function exists, that it has the above (page 58) qualitative char-

---

the process of compounding may be more complex than mere addition, because, for example, of possible reactions of the demand of individual A upon that of individual B. For purposes of the present analysis, these questions can be ignored.

[1] We remark, however, that wide diversity in the types of demand curves, for a given commodity, would ordinarily not arise in practice.

acteristics, and that its mathematical formula can take one of various assumed forms.

Thus we may assume that the collective demand function[1] is a straight line (see footnote 1, page 74)

$$p = b - \frac{b}{a} q \tag{34}$$

or a parabola

$$p = a - bq - hq^2 \tag{35}$$

or a hyperbola

$$p = \frac{k}{q} \tag{36}$$

or some more complicated form such as

$$p = \frac{m}{\sqrt{q}} \tag{37}$$

Marginal Revenue Curve. Suppose that the collective demand, for a given commodity in a stated market, is given by

$$p = b - \frac{b}{a} q \tag{34}$$

and consider a single producer, a monopolist, who is supplying the commodity for this market. An important series of problems concerns the bearing of the market demand upon the producer's output q; the q for the market is, under the suppositions of the case, controlled by this one producer.

If the producer furnishes q units to the market, the price established will be the p given by equation (34) for that q, and his total receipts, total *revenue*, will be p times q. This may be written as a new function, stating revenue R in terms of q,

$$R = pq = bq - \frac{b}{a} q^2 \tag{38}$$

Suppose that the producer is supplying q_1 units to the market for which his revenue is R_1, and then increases his output to

[1] The constants a, b, h, k in the formulas herewith which look like Equations (25), (26), and (27) are, of course, entirely new numerical values for the present case. As stated before, these collective formulas have no necessary connection with any one of the individual formulas. All the constants are positive; accordingly a constant preceded by a minus sign, such as $-h$, is a negative number.

$q_1 + \Delta q$. There will be a corresponding change, we shall call it an increase, although it may be a decrease, in which case ΔR would be negative, in R, which is designated ΔR. From Equation (38)

$$\begin{aligned} R_1 &= bq_1 - \frac{b}{a} q_1^2 \\ R_1 + \Delta R &= b(q_1 + \Delta q) - \frac{b}{a}(q_1 + \Delta q)^2 \\ &= bq_1 + b\,\Delta q - \frac{b}{a} q_1^2 - 2\frac{b}{a} q_1\,\Delta q - \frac{b}{a}(\Delta q)^2 \\ \Delta R &= b\,\Delta q - 2\frac{b}{a} q_1\,\Delta q - \frac{b}{a}(\Delta q)^2 \end{aligned}$$

and the average increase in R per unit increase in q is

$$\frac{\Delta R}{\Delta q} = b - 2\frac{b}{a} q_1 - \frac{b}{a}\Delta q$$

In the limit, as Δq approaches zero,[1]

$$\lim_{\Delta q = 0} \frac{\Delta R}{\Delta q} = b - 2\frac{b}{a} q_1 = R_m \tag{39}$$

This limiting value is called the marginal revenue for output q_1.[2]

Neglecting the subscript 1 of q, Equation (39) gives the marginal revenue as an explicit function of the quantity of output. Corresponding to this function, a curve (straight line, as it happens) can be drawn; this is shown dotted in Chart 37 (with a and b taken as 120 and 500, respectively). This marginal revenue curve is an image of the demand curve from which it is derived, also shown (solid line) in the chart.[3]

[1] Here, as before, we imagine the situation for q_1 and R_1, represented by a point A of the curve, and for $q_1 + \Delta q$ and $R_1 + \Delta R$, represented by point B. Then let B slide along the curve toward A, *i.e.*, let Δq approach zero. As the procedure is exactly similar to that followed above in connection with Chart 32 (which would serve for this purpose if we replace the variable u by R and chose the literal constants b and $-(b/a)$ as the same numbers used in Chart 32), the chart in the present case is not shown.

[2] The qualifying phrase "for output q_1" is really essential, as the marginal revenue differs for different outputs. An equivalent qualifying phrase could be written, in terms of the chart (not shown), as "at point A."

[3] This is the first case in which we have shown both a curve and its image on a single chart. As this type of construction is frequently used in economic analysis, the reader should become familiar with it.

Application of a similar procedure to the case involving a collective demand curve of the parabolic type [Equation (35)] yields

$$R = aq - bq^2 - hq^3$$

and, after calculating ΔR as the difference between $R_1 + \Delta R$ (belonging to output $q_1 + \Delta q$) and R_1 (belonging to q_1) and taking the limit of $\Delta R/\Delta q$ as Δq approaches zero,

$$\lim_{\Delta q = 0} \frac{\Delta R}{\Delta q} = a - 2bq - 3hq^2 = R_m \tag{40}$$

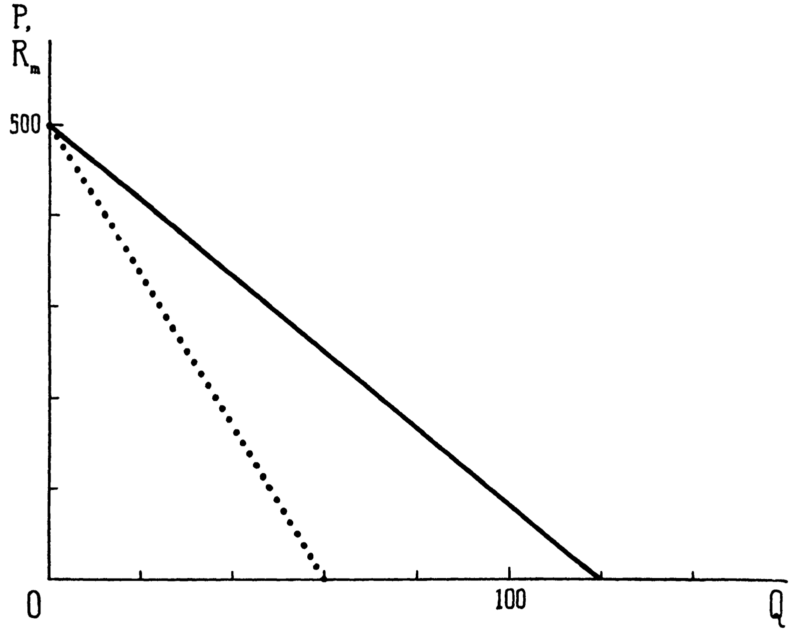

CHART 37.—Price (solid line) and marginal revenue (dotted line), for a linear-demand function.

The process of calculating the limiting value of $\Delta R/\Delta q$ used in these illustrations cannot be carried through for certain types of assumed collective demand functions. For example, if the collective demand function is

$$p = ah^{q/b} \tag{41}$$

where h is a constant (a positive fraction, less than 1) and a and b are constants, the revenue function is

$$R = aqh^{q/b}$$

and our tools from elementary algebra are not sufficiently powerful to yield a direct evaluation, by calculating the limit of $\Delta R/\Delta q$, for R_m. An attempt to apply the foregoing process leads in fact to

$$R_1 = aq_1h^{q_1/b}$$
$$R_1 + \Delta R = a(q_1 + \Delta q)h^{q_1/b+\Delta q/b}$$

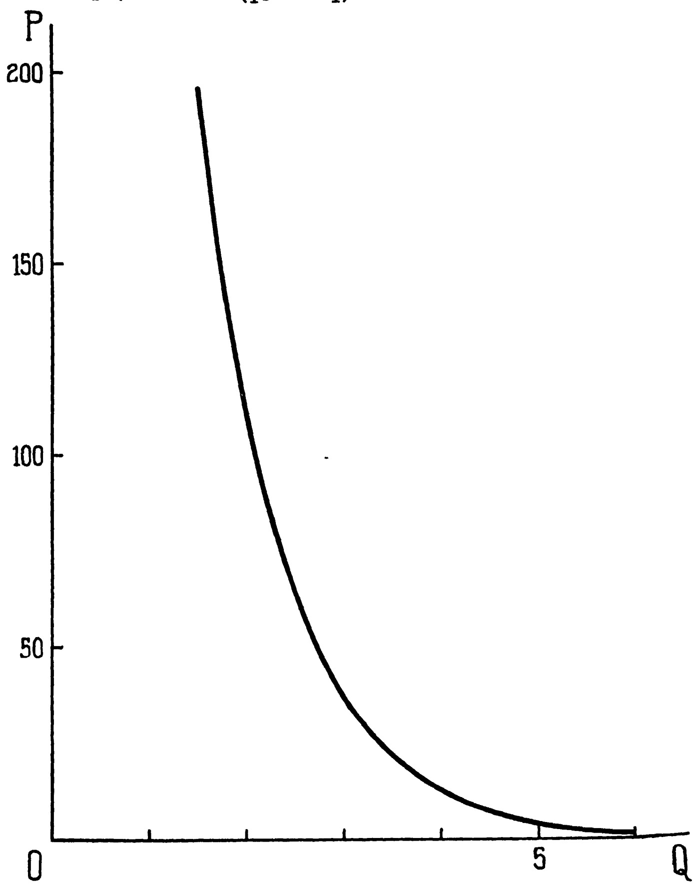

CHART 38.—Demand function of the exponential type.

which may be written

$$= aq_1h^{q_1/b}h^{\Delta q/b} + a\,\Delta qh^{q_1/b}h^{\Delta q/b}$$

giving

$$\Delta R = aq_1h^{q_1/b}(h^{\Delta q/b} - 1) + a\,\Delta qh^{q_1/b}h^{\Delta q/b}$$

and

$$\frac{\Delta R}{\Delta q} = aq_1h^{q_1/b}\frac{h^{\Delta q/b} - 1}{\Delta q} + ah^{q_1/b}h^{\Delta q/b}$$

In the limit, as Δq approaches 0, $h^{\Delta q/b}$ approaches 1, and the second term of $\Delta R/\Delta q$ approaches $ah^{q_1/b}$. But the numerator of the fraction in the first term of $\Delta R/\Delta q$ approaches $1 - 1$, which is zero, while the denominator also approaches zero. Hence, the first term is "indeterminate," so far as disclosed by the process heretofore used: it contains a fraction for which both numerator and denominator approach zero in the limit. The foregoing process thus does not readily yield the marginal revenue for a demand situation of the type reflected in Equation (41).

More generally, we assert that this direct process of evaluating the limit is laborious and involved for various types of demand functions which might be assumed and which appear tolerably in accord with the realistic characteristics of collective demand. That the method works easily for such an equation as (38) is a fortunate circumstance rather than the general rule. Treatment of certain cases, such as Equation (41), for which the direct method is not readily available is possible by methods given in the next chapter (page 94).[1]

[1] Equation (41) for h taken as $\frac{1}{5}$ (h must be a positive fraction less than 1, if the curve is to have the desired qualitative characteristics, p. 58, of a demand curve) and for a taken as 1,000 and b taken as $\frac{1}{8}$ is represented by the curve of Chart 38. (The table of values is not reproduced.) For q equal zero, p is 1,000. Mathematically, values of p exist for negative values of q, the curve sweeps upward to the left of OP, but only the portion to the right of OP has economic meaning.

CHAPTER IV

RATES AND DERIVATIVES

At numerous points in earlier chapters, interest has centered in the ratio of one variable to another. Thus, average cost of the entire output is the ratio of total cost to total output

$$\bar{c} = \frac{c}{q}$$

Similarly, average cost of an additional output Δq is the corresponding additional cost Δc divided by Δq. This latter ratio was seen to have a limiting value as Δq approaches zero.[1]

$$c_m = \lim_{\Delta q = 0} \frac{\Delta c}{\Delta q}$$

Similar instances appeared in the illustrations from demand theory, with particular reference to the utility of an individual consumer and the revenue of a monopolistic producer.

Ratios such as the foregoing may be called *rates*. They have the same generic properties as the simplest type of rates with which we are familiar, *viz.*, velocities or speeds. For example, consider a motor journey from a stated point in New York City to a stated point in Boston; let the total distance covered be 239 miles, and the time be 6½ hours. The rate at which the distance was covered would be distance divided by time, 239 ÷ 6.5, which is approximately 36.7 miles per hour. Such a rate is sometimes distinguished from other rates by calling it a *time rate*, the word "time" specifying the nature of the independent variable. It is also called a *velocity* (although

[1] We ignore here the possibility that, in the case of certain types of functions, the limit may not exist or may be indeterminate for certain values of q. The curves studied above were purposely selected to avoid troublesome instances of this kind. For the present, the student need not be concerned about cases for which the limit does not exist, for they do not ordinarily arise in elementary economic problems.

velocity comprises direction of movement as well as rate) or speed.[1]

Average and Instantaneous Rates. The most significant fact about the rate of 36.7 miles per hour, discussed above, is that it is an average rate. It implies that, if the motor had been driven at a constant velocity throughout the trip and if that velocity had been 36.7, the total distance would have been covered in the same time as actually taken. The obvious fact of experience, however, is that no motor could make such a trip at constant velocity: not only would the actual velocity vary over different portions of the journey, but stops along the way would presumably occur, and at such times the true velocity would be zero. All this is well known to the motorist; he is fully aware that, in order to have an average speed as high as 36.7 between New York and Boston, he must attain an actual speed much above 36.7 on some parts of the journey.

The average rate is thus a resultant, or composite, of the variable actual rates prevailing along different parts of the route. Assuming that the motor's speedometer was not functioning, the driver might seek to determine approximately his actual rate for one or more portions of the journey. Thus, he might observe the time taken between two towns along the route, and by taking the map distance between the towns, calculate the ratio of distance divided by time. The result would be the time rate between those two towns; but it would still be an average rate, for the actual speed might vary widely along the way between the towns. Or, he might clock the time taken over one of the 1-mile stretches marked off at certain places on main highways, divide unity (1 mile) by the time noted, and have his speed—still an average rate—for that particular marked mile. Neither of these procedures gives him the actual speed at any particular point of the trip; although, by taking sufficiently frequent observations for a succession of sufficiently short distances, he can get as near as he wishes to a complete tabulation of his varying speed. And yet, even though he broke the entire distance into a very large number of very small segments, which amounts to breaking the entire time of the trip into many very

[1] Thus, by definition, speed is a rate, the rate at which distance is traveled per unit time. The colloquial expression "rate of speed" is therefore redundant and should always be shunned in precise speaking.

small intervals, and calculated the average speed for each, he would still not know the actual speed at any point, waiving the fact that he knows the speed is zero whenever he is standing still. The speed, at a particular time of the trip, which the motorist seeks to measure by calculating the average speeds for briefer and briefer intervals following that moment—the actual speed at a particular point of the route, or at a particular instant of time—is called an *instantaneous rate.*

The student will readily recognize the limit notion in the concept of an instantaneous rate. We may in fact define the instantaneous rate at a particular instant of time as the limit of the average rate during an interval of time immediately following that instant, as the said interval approaches zero as a limit. Accordingly, the attempt of the motorist, described above, to determine his actual rates at various points along the route by breaking the total distance into numerous small segments and calculating the average rate for each segment yields an approximation to the instantaneous rate at any point within each segment. As the segment becomes shorter and shorter, the time interval approaches zero, the average rate in the segment approaches the instantaneous rate as a limit.[1]

Although our early contacts with the rate concept run in terms of time rates, distance divided by time, the concept is much more general. Some variable other than time can be the independent variable in a rate, the ratio of a dependent variable to such independent variable is just as truly a rate. The student should therefore break the habit of thinking of a rate as a time rate, a speed, or velocity, and come to think of it in the general sense: a *rate* is a ratio that describes the variation of one variable with reference to the variation in another variable. For all rates of this more general sort, the distinction between average rate and instanteous rate remains of basic importance.

Thus the average cost for a specified output q

$$\bar{c} = \frac{c}{q}$$

[1] The speedometer is a mechanical device for determining approximately the instantaneous rate, although it does not depend essentially upon the limit principle; but technical imperfections result in its yielding in fact an average rate for a very brief interval, rather than the true instantaneous rate. Some speedometers are in fact slow to respond to changes in actual speed.

is an average rate. It simply means that, if the cost per unit were constant (fixed for all units of output, or more strictly for each portion of the output), the total cost for output q would be c. Likewise, if the output at the scale q is increased by an amount Δq, with a corresponding increase Δc in cost, the average cost of the additional output

$$\frac{\Delta c}{\Delta q}$$

is an average rate. It merely implies that, if the rate of change in cost relative to quantity were fixed at the level $\Delta c/\Delta q$, an increase Δq in output would bring an increase Δc in cost.[1]

On the other hand, the marginal cost

$$c_m = \lim_{\Delta q = 0} \frac{\Delta c}{\Delta q}$$

is an instantaneous rate. Like the speed indicated by the speedometer of an automobile, it pertains to a specific point and not to an interval. In the automobile case, the interval is an interval of time, but here it is an interval of quantity: the

[1] In these instances, when we talk of a "change" or an "increase," we feel an impelling inclination to think of such change or increase as a definite movement, which therefore involves the notion of time. This insidious entrance of the time element into all our thinking about variation must be resisted: we must come to think of a variation as a mere difference, with no implication that a change takes place and requires a lapse of time in which to take place. Variations, in the sense of differences, need contain no time element.

Variations, in the sense of "differences" not involving a time element, may be illustrated by the temperatures at various points in a room. If we observed the temperatures at various points, ranging from a point near the radiator to a point near a window exposed to a wintry wind, the observations would surely record variation. If we used a single thermometer, there would be an "insidious entrance of the time element": time would actually be consumed in moving the thermometer from one point to another, and the observations would be ordered successively in time. If, however, we used numerous thermometers, all identically calibrated, and stationed at the various points with enough observers to read them simultaneously, variation would still be recorded; but no time element would be present, and the variation would involve mere differences. This confusion in the sense of the word variation, between the sense of a course of change over time and the sense of mere simultaneous differences, often perplexes the student of economics, and he must learn early to make the necessary distinction.

distinction relates merely to a difference in the nature of the independent variable.

Likewise, all the marginal concepts discussed in the previous chapter are instantaneous rates. Marginal utility is the instantaneous rate of change of utility with quantity. Marginal revenue

$$R_m = \lim_{\Delta q = 0} \frac{\Delta R}{\Delta q}$$

is the instantaneous rate of change of revenue with quantity.

It so happens that the economic cases of rates so far considered include only instances in which quantity is the independent variable, but numerous instances having other independent variables can arise in practice. The concepts of average rate and instantaneous rate—the word "instantaneous" meaning "at a point" and not "at an instant of time"—and the analytical relationships presently to be developed concerning those concepts are perfectly general. They are in no way limited to such a specific independent variable as time or as quantity.

Rate of Profit. For example, the same concepts appear in studying the return (or profit) on capital investment. Suppose that, for a stated time interval such as a year, the aggregate net profit—defined in some appropriate way, such as the remainder of income after all operating costs like wages and expenditures for materials and all fixed costs like interest and taxes are deducted—is P_1 for a specified industrial enterprise with the amount of capital invested fixed during the interval at I_1.[1] The average return on capital, the *average rate of profit*, is then

$$\bar{P}_1 = \frac{P_1}{I_1}$$

[1] In an actual enterprise, the capital normally does not remain constant during an interval even so brief as a year. For the present purpose, this difficulty is waived by assuming capital fixed at its average for the year. We are in truth dealing here with magnitudes that are themselves rates, average time rates. Thus, in fact, P_1 is the amount of profit per year for the year under examination; the phrase "per year" implies an average time rate. Of necessity, many economic magnitudes are such time rates because we are obliged to make measurements in terms of an interval of time. But this entrance of the rate notion into the very definition of our economic magnitudes can be ignored in considering the very different type of rate here under study (see foonote, p. 1).

With the enterprise as it actually exists, the entire capital investment is a single entity; and to think of the possibility that some portion of it, say a particular machine, or a particular pecuniary amount such as $1,000, may yield a profit at a rate differing from $\bar{P}_1$ appears somewhat artificial.

Nevertheless, we can helpfully examine the rate of variation of profit with investment. We may imagine that investment is somewhat different from I_1, say I_2, and that correspondingly profit amounts to P_2 rather than P_1. For this second assumed situation, average rate of profit is

$$\bar{P}_2 = \frac{P_2}{I_2}$$

Manifestly, we can conceive of profit as a function of investment, just as cost was regarded (Chap. I) as a function of output. Such a functional relation between profit and investment shows, symbolically, or graphically if it is charted, how profit varies with investment. Again, the student is cautioned against reading a time element into the word "varies": we are thinking here merely of different amounts of profit associated with different amounts of investment, and all these various amounts are imagined as applicable to a single instant, or interval, such as a stated year, of time.[1]

By the customary procedure followed above, let

$$\begin{aligned} I_2 &= I_1 + \Delta I \\ P_2 &= P_1 + \Delta P \end{aligned}$$

and then represent the average rate of profit on the additional investment by

$$\frac{\Delta P}{\Delta I}$$

[1] As in previous functional relations studied, cost associated with quantity, the manner of discovery of the function here may be obscure. In the present instance, as in the others, the alternatives are the a priori and the empirical methods; but practical obstacles in the way of the second approach are even more evident than in the case of the individual demand curve (see p. 59). One of the chief obstacles arises because only one experiment can be tried upon a given enterprise at a given time.

By letting ΔI approach zero as a limit,[1] we have the marginal, or instantaneous, rate of profit

$$P_m = \lim_{\Delta I = 0} \frac{\Delta P}{\Delta I}$$

The Derivative Concept. Instantaneous rates, of the sort illustrated by such marginal magnitudes as marginal cost, marginal utility, and marginal return on investment, have roles so extensive in various fields of applied mathematics that a special methodology has been evolved for their symbolic treatment. This methodology is called *differential calculus.* Such a methodology, while it involves technical and theoretical problems interesting for their own sake, may be regarded by the economist as a mere working tool that facilitates the application of the rate notion, and the limit notion, to economic reasoning. In this sense, it is like any other timesaving technical device, such as a multiplication table, which presents the products, worked out once for all, of various pairs of numbers that would otherwise need be multiplied every time one needed their products.

The difference from such a simple timesaving device as the multiplication table is one of complexity. Differential calculus is sufficiently complicated, in its definitions and implications, so that the user of the method must be more carefully informed than when using a mere multiplication table. It is a mistake, however, to dread some deep mystery in the calculus: the root ideas are not mysterious and the method that rests upon them is wholly logical and readily understandable to anyone who has disciplined his mind to precise thinking. If the reader has carefully mastered the foregoing treatment of limits, he need encounter no serious difficulty in grasping the significance of those elementary calculus notions which are pertinent to the simpler mathematical problems in economics. He will presently find himself possessed of a powerful tool of analysis, which will not only save time in the study of relationships among economic variables, but also promote precision of reasoning in such study.

[1] Again, we waive the possibility that the limit may not actually exist for one or more points on the curve relating profit to investment. We assume for the present, though this may be at variance with practical fact under certain conditions, that the curve has a simple form for which the limit does exist at every point.

The basic concept in the differential calculus is that of the *derivative*—more strictly expressed, the "derivative of a dependent variable y with respect to an independent variable x"—and it is generally represented by the compound symbol[1]

$$\frac{dy}{dx}$$

A possible, and satisfactory, definition of the derivative, in the light of the foregoing discussion of rates, would read: The derivative of y with respect to x is the instantaneous rate of change of y with x. More generally, however, the derivative is defined directly in terms of the limit concept, which, as observed above, is involved in the notion of instantaneous rate.

Thus, if y is a function of x satisfying certain rather liberal requirements,[2] we consider the situation for some specified value x_1 of x and then consider a change Δx in x, accompanied by a change Δy in y. In the jargon of the calculus, these changes, Δx and Δy, are generally called *increments*. The ratio $\Delta y/\Delta x$ is the average rate of increase in y as x increases from x_1 to $x_1 + \Delta x$. The derivative of y with respect to x, for x equal to x_1,[3] is defined as the limit of this ratio, as Δx approaches zero; thus, symbolically,

$$\frac{dy}{dx} = \lim_{\Delta x = 0} \frac{\Delta y}{\Delta x} \tag{42}$$

[1] It is in the highest degree important that the student never forget that this is a compound symbol; it is a single symbol meaning "the derivative of y with respect to x." It is not dy "divided by" dx; the upper portion is not d "times" y; and the lower portion is not d "times" x. Mathematicians do in fact treat dy/dx in some respects and for some symbolic operations as if it were dy "divided by" dx; the student may already have encountered examples of such treatment (see p. 151). In all such cases, however, the careful worker never forgets the true nature of the compound symbol. Anyway, the student who is beginning his excursion into the unknown realm of calculus will do well to adhere strictly to the basic statement: dy/dx is a single symbol; and it should not be tampered with, for example, by treating it as a fraction.

[2] In most elementary economic problems, these requirements are met and need not concern us.

[3] Strictly, this phrase is essential, although in practice it is frequently omitted from the definition. Its presence serves to remind us that the derivative may, and probably does, have different values as x changes. For simple functions, such as those studied above, the limit has a value, and thus the derivative exists, for each value of x_1.

The foregoing definition of the derivative, identifying it with the limit of a ratio established in a specified manner, is perfectly general. Every case discussed above, of a marginal magnitude defined as such a limit, therefore involves a derivative. Thus, marginal cost is the derivative of cost with respect to output

$$c_m = \frac{dc}{dq} = \lim_{\Delta q = 0} \frac{\Delta c}{\Delta q}$$

marginal utility is the derivative of utility with respect to quantity demanded

$$u_m = \frac{du}{dq} = \lim_{\Delta q = 0} \frac{\Delta u}{\Delta q}$$

and marginal return on investment is the derivative of profit with respect to investment

$$P_m = \frac{dP}{dI} = \lim_{\Delta I = 0} \frac{\Delta P}{\Delta I}$$

Likewise, other marginal concepts, besides those discussed above, can, in all cases where the marginal concept implies the limit of a ratio of the type shown, be called derivatives. Moreover, even though the term "marginal" is not attached to the concept, so long as the limit of a ratio, of the instantaneous-rate sort, is involved, the concept can clearly be represented by a derivative. Consequently, if we can build up a scheme of analysis for manipulating the derivative, subjecting it to mathematical operations of various sorts, and exploring its relationships to the variables x and y and functions of x and y (expressions containing them), we can at once use this scheme as a general tool for studying functions of the marginal, or instantaneous-rate, type.

Differentiation. The foregoing definition of derivative implies that y is a function of x; y may be any dependent variable, such as cost, and x may be a corresponding independent variable, such as quantity of output. The first step, in developing a technical scheme for analyzing derivatives and studying problems that involve derivatives, is the discovery of methods for determining the derivative for any particular function. Such a method, called *differentiation*, may be expected to yield promptly the derivative in any case as soon as the functional relation between

y and x is known,[1] just as a logarithm table yields the logarithm of a number as soon as the number is known. The formula, which tells what the derivative is for any specified function, is obtained basically by a direct application of the limit principle involved in the definition of derivative. In other words, these formulas are found by applying the definition of derivative. Once the formula has thus been worked out, it can thereafter be used, without resort to the laborious limit calculation, as a guide for writing down the derivative immediately by inspection.[2]

Suppose, for example, the known function is a positive integral power of x (say, the fourth power)

$$y = x^4$$

Following the procedure of an earlier chapter, for finding the limit of $\Delta y/\Delta x$ as Δx approaches zero

$$\begin{aligned}
y_1 &= x_1^4 \\
y_1 + \Delta y &= (x_1 + \Delta x)^4 \\
&= x_1^4 + 4x_1^3\,\Delta x + 6x_1^2(\Delta x)^2 + 4x_1(\Delta x)^3 + (\Delta x)^4 \\
\Delta y &= 4x_1^3\,\Delta x + 6x_1^2(\Delta x)^2 + 4x_1(\Delta x)^3 + (\Delta x)^4 \\
\frac{\Delta y}{\Delta x} &= 4x_1^3 + 6x_1^2\,\Delta x + 4x_1(\Delta x)^2 + (\Delta x)^3
\end{aligned}$$

and hence

$$\frac{dy}{dx} = \lim_{\Delta x = 0} \frac{\Delta y}{\Delta x} = 4x_1^3$$

To make this a general formula, for any value of x as well as for x_1, we drop the subscript 1 and have

$$\frac{dy}{dx} = 4x^3 \tag{43}$$

[1] Actually, for cases encountered in elementary economics, we need derivatives for only a few fairly simple types of functions. The methods outlined below, therefore, cover only a limited range of functions for which derivatives might be desired in various branches of applied mathematics. (Some are included below which are not needed in elementary economics.)

[2] Just as, in elementary algebra, we have the formula

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

which enables us to write down at once

$$(2 + 3k)^3 = 8 + 36k + 54k^2 + 27k^3$$

instead of going through the lengthy work of multiplying $(2 + 3k)$ by $(2 + 3k)$ and then multiplying the result by $(2 + 3k)$.

This formula can always be used as a guide for writing down the derivative for this particular function, x^4. The formula can be memorized, or it can be preserved in some accessible list or index, where it can be found when needed. In any case, as soon as we know that

$$y = x^4$$

we also know that

$$\frac{dy}{dx} = 4x^3$$

This formula pertains to the particular case for y equal to the fourth power of x. We observe that the derivative in this case takes the form of "4 times the third power of x." This suggests the rule: The derivative with respect to x of the kth power of x is equal to k (where k is a positive integer) times the $(k - 1)$st power of x. This rule can in fact be proved, by a process of computing the limit wholly analogous to the above, for any k that is an integer, positive or negative. The proof is here omitted, but the resultant more general formula is now given for

$$y = x^k$$

$$\frac{dy}{dx} = kx^{k-1} \tag{44}$$

By somewhat more complicated analysis, also omitted here, this result can be established also for fractional values, positive or negative, of k. Equation (44) therefore holds for every k. Thus

For	k is	Derivative is
$y = x^2$	2	$2x$
$y = x^{-5}$	-5	$-5x^{-6}$
$y = x^{2/3}$	$2/3$	$2/3\, x^{-1/3}$
$y = x^{-6/7}$	$-6/7$	$-6/7\, x^{-13/7}$

Differentiation Rules. Formulas like Equation (44) can be worked out for various types of function; each such formula tells what the derivative of y with respect to x will be when y is the stated function of x. Such formulas are rules for differentiation. These rules can be systematically used for obtaining the derivative for any stated function; knowledge of the rules, or an accessible list of them, enables the student to write down

at sight the derivatives of those functions commonly encountered in economics.[1]

Proofs of the various rules of differentiation, corresponding to the proof of Equation (43), will not be given here; they can be found in various elementary texts on calculus. Broadly speaking, the proofs rest upon the same straightforward application of the limit notion, *i.e.*, the direct application of the definition of derivative, as used in deriving Equation (43); but, for certain functions, an indirect procedure in proof is advantageous or necessary. For several cases presented in the rules below, proof of the particular formula follows readily from the general formula for a function of a function, Rule 14. Even in cases of indirect proof, however, the limit concept involved in the definition of the derivative is the ultimate foundation of the proof; the reader may fairly assume that all differentiation rules depend, directly or indirectly, upon the limit concept. In order to have important differentiation rules available for use in the following and other analyses of economic problems, a partial list of such differentiation rules is now presented, without proof.

1. The derivative of a constant, with respect to any variable, is zero. Thus, if k is a constant,

$$\frac{dk}{dx} = 0$$

2. The derivative of a constant multiplied by a function is the constant multiplied by the derivative of the function. Thus, if k is a constant and v is a function of x,

$$\frac{d(kv)}{dx} = k\frac{dv}{dx}$$

3. The derivative of the sum of two (or more) functions is the sum of the derivatives of the separate functions. Thus, if v and w are functions of x,

$$\frac{d(v + w)}{dx} = \frac{dv}{dx} + \frac{dw}{dx}$$

[1] Nothing is said here about functions that have no derivative, as defined, for certain values of x, or in some cases, for any value of x. Such functions do not ordinarily arise in economics, and the student may for the present safely assume that every economic function does have a derivative for each value of x.

4. The derivative of the product of two functions is the sum of two terms, the first function multiplied by the derivative of the second, and the second function multiplied by the derivative of the first. Thus, if v and w are functions of x,

$$\frac{d(vw)}{dx} = v\frac{dw}{dx} + w\frac{dv}{dx}$$

4*a*. The derivative of a quotient is the derivative of the numerator, divided by the denominator, minus the derivative of the denominator, multiplied by the numerator and divided by the square of the denominator,

$$\frac{d\left(\frac{v}{w}\right)}{dx} = \frac{1}{w}\frac{dv}{dx} - \frac{v}{w^2}\frac{dw}{dx}$$

5. The product of several functions yields a derivative made up of several terms, each term being the derivative of one function multiplied by the other functions. Thus, if t, u, v, and w are functions of x,

$$\frac{d(tuvw)}{dx} = uvw\frac{dt}{dx} + tvw\frac{du}{dx} + tuw\frac{dv}{dx} + tuv\frac{dw}{dx}$$

A convenient alternative handling of this case is to note that $\log (tuvw) = \log t + \log u + \log v + \log w$, and then use rule 9.[1]

6. The derivative, with respect to x, of any constant power of x, where the power is designated by the (constant) exponent a, is a multiplied by a power of x having an exponent less than a by unity. Thus

$$\frac{dx^a}{dx} = ax^{a-1}$$

[This is the rule already given in Equation (44).]

7. For the constant power of a function of x, the rule becomes

$$\frac{dv^a}{dx} = av^{a-1}\frac{dv}{dx}$$

where v is a function of x.

[1] This expression for the logarithm of a product is explained, as a direct consequence of the definition of a logarithm, in texts on college algebra (see footnote 1, p. 73).

8. The derivative, with respect to x, of the logarithm of x (the logarithm to the Napierian base[1]) is unity divided by x. Thus

$$\frac{d \log x}{dx} = \frac{1}{x}$$

9. More generally, for the logarithm of a function v of x, the rule becomes

$$\frac{d \log v}{dx} = \frac{1}{v}\frac{dv}{dx}$$

where v is a function of x.

10. The derivative of the logarithm, to some base other than e, such as b, of a function v of x is given by the rule[2]

$$\frac{d \log_b v}{dx} = (\log_b e)\frac{1}{v}\frac{dv}{dx}$$

11. The derivative of a variable power of the Napierian constant e, represented by the variable exponent v, where v is a function of x, is

$$\frac{de^v}{dx} = e^v\frac{dv}{dx}$$

11*a* and 11*b*. As special cases of rule 11

$$\frac{de^x}{dx} = e^x$$

and, where k is a constant,

$$\frac{de^{kx}}{dx} = ke^{kx}$$

12. The derivative of a variable power of any constant b, represented by the variable exponent v, where v is a function of x, is[3]

$$\frac{db^v}{dx} = (\log_e b)b^v\frac{dv}{dx}$$

[1] The Napierian, or "natural" base of logarithms is approximately 2.71828183, usually designated by e. The symbol "log," as used in theoretical analysis, customarily implies the base e; for any other base, such as b, the symbol is $\log_b$; but, by exception, the base 10, frequently used in computation by logarithms, is often not explicitly attached to the log symbol (see footnote 1, p. 73).

[2] In rules 10, 12, 12*a*, and 12*b*, the symbol $\log_b e$ (or $\log_e b$) is an entity by itself; the log operation does not "cover" anything after the e (or b).

[3] See preceding footnote.

12*a* and 12*b*. As special cases of rule 12:

$$\frac{db^x}{dx} = (\log_e b)b^x$$

and

$$\frac{db^{kx}}{dx} = k(\log_e b)b^{kx}$$

13. The derivative of a *variable* power of a *function* w, represented by the variable exponent v, where v is also a function of x, is

$$\frac{dw^v}{dx} = vw^{v-1}\frac{dw}{dx} + (\log_e w)w^v\frac{dv}{dx}$$

14. The derivative of a function w of v, where v is a function of x with respect to x, is

$$\frac{dw(v)}{dx} = \frac{dw}{dv}\frac{dv}{dx}$$

15. The derivatives of the simple trigonometric functions of v, where v is a function of x, are

$$\frac{d \sin v}{dx} = \cos v \frac{dv}{dx}$$

$$\frac{d \cos v}{dx} = -\sin v \frac{dv}{dx}$$

$$\frac{d \tan v}{dx} = \sec^2 v \frac{dv}{dx}$$

$$\frac{d \cot v}{dx} = -\csc^2 v \frac{dv}{dx}$$

$$\frac{d \sin^{-1} v}{dx} = \frac{1}{\sqrt{1 - v^2}}\frac{dv}{dx}$$

$$\frac{d \cos^{-1} v}{dx} = -\frac{1}{\sqrt{1 - v^2}}\frac{dv}{dx}$$

$$\frac{d \tan^{-1} v}{dx} = \frac{1}{1 + v^2}\frac{dv}{dx}$$

$$\frac{d \cot^{-1} v}{dx} = \frac{-1}{1 + v^2}\frac{dv}{dx}$$

where the superscript -1 on the left side of the last formula means "the angle whose cotangent is" v, and analogously for the three preceding formulas. These four functions, the anti-trigonometric functions, are not single-valued, but multiple-valued; the formulas here given apply only to a single such value.

Not all the foregoing rules of differentiation will be needed in elementary economic work; but they are brought together here for convenience and cover adequately all cases of differentiation in elementary mathematics. Rules for still more intricate functions can be found in treatises on calculus; there also are discussed functions that have no derivatives, at some or all points (for some or all values of x).

Examples. Facility in the use of these rules will come only with practice, and the reader should accustom himself to writing down, by the use of appropriate rule or rules, the derivative of any particular function encountered. That more than one rule may need be applied in a particular case should be kept constantly in mind, although many simple cases are covered by a single rule.

Thus, for the function given by Equation (16) of Chap. III,

$$c = 0.2q^2 + 1.5q + 7$$

rules 3, 2, 6, and 1 apply, giving

$$c_m = \frac{dc}{dq} = 0.4q + 1.5$$

This is precisely the result previously obtained, in Equation (24), by the laborious direct application of the limit notion.

Similarly, for Equation (28) of Chap. III,

$$u = bq - \frac{b}{a}q^2$$

rules 3, 2, and 6 apply, giving

$$u_m = \frac{du}{dq} = b - 2\frac{b}{a}q$$

as found, by the laborious method, in Equation (30).

These illustrations bring out the general fact that the rules of differentiation can be used as a direct means of writing down the limit, as soon as the basic function is known, in all cases

similar to those discussed in Chap. III which involve the marginal, *i.e.*, the limit or instantaneous-rate, concept. In those cases, and similar cases, differentiation may for present purposes be regarded as a mere tool for furnishing by inspection the result that would otherwise need be calculated by the laborious direct application of the limit process (as at page 55). The following chapters will develop further uses of the derivative—uses in which important analytical operations not heretofore discussed will be found facilitated by the derivative concept. These and other analytical uses of the derivative, rather than its mere equivalence to the limit or instantaneous-rate notion, constitute its great power in applied mathematics. In these uses the derivative frequently enables us to perform symbolic operations that would otherwise fall beyond our capacity for analysis.

Successive Differentiation. Before proceeding to these more elaborate applications of the derivative, certain further technical points about derivatives can helpfully be developed. Once the derivative of a function has been found, it can itself be considered as a new function. Thus, if the given function is

$$c = aq^2 + bq + k \tag{45}$$

its derivative (with respect to q) is

$$\frac{dc}{dq} = 2aq + b$$

and this is a new function, which we have called c_m.

Now, the function

$$c_m = 2aq + b \tag{46}$$

also has a derivative (with respect to q), and this derivative can be written down by inspection from the rules of differentiation. Thus,

$$\frac{dc_m}{dq} = 2a \tag{47}$$

Just as Equation (46) gives the marginal function related to the function of Equation (45), so Equation (47) gives the marginal function related to the function of Equation (46). Just as Equation (46) gives the instantaneous rate of change, in terms of q, of the function c defined by (45), so Equation (47) gives the instantaneous rate of change of c_m as defined by (46).

In arriving at Equation (47), we could consider Equation (46) as giving a new or isolated function c_m, forgetting that c_m is itself the derivative of some other function c, given by (46). But, remembering that c_m is the derivative of c, we may appropriately say that (47) gives the "derivative of a derivative"; (47) gives the derivative of a function which is itself the derivative of some other function. For this reason we are justified in calling the right side of Equation (47) the "second derivative of c with respect to q," and it is customarily represented by the symbol[1]

$$\frac{d^2c}{dq^2}$$

The reader will observe that the second derivative of a function, with respect to an independent variable, can be obtained merely by writing down the derivative (now advantageously called the first derivative) of the function, by use of the rules of differentiation, and then writing down the derivative of the first derivative. This process is called *successive differentiation.* Thus, if

$$y = 8x^4 - 6x^2 + 3x - 9 \tag{48}$$

the second derivative, of y with respect to x, is obtained as follows:

$$\frac{dy}{dx} = 32x^3 - 12x + 3$$

and

$$\frac{d^2y}{dx^2} = 96x^2 - 12$$

Moreover, this procedure can be followed no matter how complicated the original function, so long as it is covered by the known rules of differentiation. Thus, for

[1] This is also a compound symbol; strictly, it should be read "the second derivative of c with respect to q." It certainly should not be read: "d squared times c divided by d times q squared"—this statement would be completely wrong as a whole and in all its parts. The symbol is not a fraction, and the operations of "squaring" and "multiplying" are not involved, either in the upper line or in the lower. Mathematicians sometimes call it loosely "d second c dq second"; but the beginner will do well to adhere rigidly to the reading "the second derivative of c with respect to q," and thus constantly remind himself of what the symbol precisely means.

$$y = \log (2x^2 + 9)$$
$$\frac{dy}{dx} = \frac{1}{2x^2 + 9} 4x$$
$$\frac{d^2y}{dx^2} = \frac{-1}{(2x^2 + 9)^2} 16x^2 + \frac{4}{2x^2 + 9} = \frac{-8x^2 + 36}{(2x^2 + 9)^2}$$

Successive differentiation can manifestly be carried beyond the second stage, and subject to our ability to calculate the successive derivatives by using the rules of differentiation or otherwise, derivatives of any order can be found. The first derivative is called that of the *first order*, the second derivative, that of the *second order*, and so on. Thus, for the above function

$$y = 8x^4 - 6x^2 + 3x - 9 \tag{48}$$

derivatives of the first five orders are[1]

$$\frac{dy}{dx} = 32x^3 - 12x + 3$$
$$\frac{d^2y}{dx^2} = 96x^2 - 12$$
$$\frac{d^3y}{dx^3} = 192x$$
$$\frac{d^4y}{dx^4} = 192$$
$$\frac{d^5y}{dx^5} = 0$$

The Expansion of a Function of One Variable. The notion we have just acquired of derivatives of second, third, . . . order, or "higher derivatives," will now be applied in order to display a device that is as important in economics as it is in physics. For this purpose it is convenient to introduce a different notation. When we wish to express the fact that a functional relation exists between a dependent variable y and an independent variable x, we simply write (see page 7 and pages 68 and 89)

$$y = f(x)$$

[1] An interesting fact appears in the final equation: whenever a function, such as Equation (48), consists solely of the sum of positive integral powers of x (each possibly multiplied by a constant), successive differentiation can be carried to a point where the derivative of a certain order, the order one greater than the highest power of x, is zero.

If such a function possesses derivatives of the first, second, third, . . . order, we denote them by $f'(x)$, $f''(x)$, $f'''(x)$, . . . , so that

$$\frac{dy}{dx} = f'(x), \qquad \frac{d^2y}{dx^2} = f''(x), \qquad \frac{d^3y}{dx^3} = f'''(x), \text{ etc.}$$

Let us return to the straight-line demand law already discussed (where the present b is $-a$ divided by the old b) (page 61, Chart 28) and rewrite it in the form

$$q = f(p) = a + bp, \qquad \text{where } a > 0,\ b < 0$$

Then,

$$\frac{dq}{dp} = f'(p) = b$$

Now we choose a particular point (p_0, q_0) that lies on that straight line. If we increase the price p_0 by a small amount, say h, we may represent the changed quantity sold, changed because of this price increase, by

$$f(p_0 + h) = a + b(p_0 + h) = f(p_0) + hf'(p_0)$$

Manifestly, this result will not hold for a nonlinear demand law. For instance, in the case of a parabolic law

$$q = f(p) = a + bp + cp^2, \qquad \text{where } a > 0,\ b < 0,\ c > 0$$

we find that

$$f'(p) = \frac{dq}{dp} = b + 2cp, \qquad \text{and} \qquad f''(p) = \frac{d^2q}{dp^2} = 2c$$

Again let us fix on any point whose coordinates (p_0, q_0) satisfy the demand law under consideration, and see how the quantity q_0 is affected by a small increase h of the price. We have

$$\begin{aligned} f(p_0 + h) &= a + b(p_0 + h) + c(p_0 + h)^2 \\ &= a + bp_0 + cp_0^2 + h(b + 2cp_0) + h^2c \\ &= f(p_0) + hf'(p_0) + \frac{h^2}{2} f''(p_0) \end{aligned}$$

The student will observe that, although we have now arrived at a result that differs from the one obtained in the case of the linear demand function, there is nevertheless a suggestive family likeness between the two. The question may arise in his mind whether all types of functions can be expressed in this con-

venient way, simply by adding a sufficient number of terms of similar form, to wit

$$f(p_0 + h) = f(p_0) + \frac{h}{1} f'(p_0) + \frac{h^2}{1 \cdot 2} f''(p_0) + \frac{h^3}{1 \cdot 2 \cdot 3} f'''(p_0) + \cdots + \frac{h^n}{1 \cdot 2 \cdot 3 \cdots n} f^n(p_0)$$

For many types of functions the answer is negative. Only a restricted class of functions can be so expressed. All the more important is it that, among all those functions which cannot be so *expressed*, there is a very large class that can be *approximated* in this manner. This means two things: (1) by adding more and more terms of the form

$$\frac{h^n}{1 \cdot 2 \cdot 3 \cdots n} f^n(p_0), \qquad n = 1, 2, 3, \text{etc.}$$

we can, in the case of this large, but limited, class of functions, never reach the exact value of $f(p_0 + h)$; but (2) as we go on adding further terms, or, as we go on increasing the number n, we get nearer and nearer to the exact value.[1] The series, known as "Taylor's series,"

$$f(p_0 + h) = f(p_0) + \frac{h}{1} f'(p_0) + \frac{h^2}{1 \cdot 2} f''(p_0) + \frac{h^3}{1 \cdot 2 \cdot 3} f'''(p_0) + \cdots \quad (49)$$

[1] An example will help the student to grasp the essential point. Consider the sum

$$S = 1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{16} + \cdots$$

It is easy to see that, as we go on adding further terms of the same type, S steadily approaches the number 2, though it never reaches 2 however many terms we add. Therefore this series is called an *infinite series* that "converges" to 2. The five terms written down already take us to $3\frac{1}{16}$. If we stop at that, we are indeed neglecting an "infinite number" of terms. But the sum of this infinite number of terms amounts to only $\frac{1}{32}$ of the value we aim at, and for many purposes, especially in economics, an "error" of this order of magnitude need not cause concern. The general criterion of convergence in this sense may be stated as follows.

We call a series convergent if, having chosen a certain term, say the mth, we find that the k terms subsequent to the mth [the $(m + 1)$st, $(m + 2)$ . . .], however large we may make the k, sum up to no more than a number ϵ, which is small enough to be neglected. How small the ϵ actually has to be in any individual case depends, of course, on the purpose in hand and also on the standards of the investigator.

can be extended indefinitely, by adding further terms; it is an infinite series of terms. But we may stop after a number n of terms and neglect the *remainder*, provided we are content with the approximation thus attained. We cannot here go into the conditions that a function must fulfill in order to be amenable to approximate expression, "expansion," by a Taylor series. This method of approximating the value of a function in the neighborhood of a given point is useful only if the number of terms required for a reasonable approximation is not too great. In practice, economists as well as physicists have developed a habit of stopping at the term containing the first derivative and simply writing, as approximately correct,

$$f(p_0 + h) = f(p_0) + hf'(p_0)$$

This procedure which is known as *expanding a function and neglecting second and higher powers* (*of h*) of course presupposes not only that the function in question possesses a Taylor series that expresses it and is convergent,[1] but also that this Taylor series converges very rapidly. The student will perceive, however, that the risk involved is greatly reduced if h is very small. If, for instance, we are investigating the effect upon quantity demanded of a small increase in price, such as may occur in response to the imposition of a small tax per unit of the commodity under consideration, we are presumably within our rights if we equate that effect to the small increase in price multiplied by the value, taken at the old price, of the derivative of quantity with respect to price. Whenever the demand function is not linear, this can be only approximately true. But in all ordinary circumstances it will be approximately true, if the increase in price is small enough, though not otherwise. This fact is important to remember because of the serious limitations it imposes upon the applicability of many propositions of economic theory. Taxes, for instance, are not usually "very small" in real life.

Differentiation with Respect to Several Variables. A further type of differentiation, somewhat analogous to successive differentiation with respect to a single independent variable, is of importance for certain problems in applied mathematics (including some problems in economic theory) in which the

[1] See preceding note.

given dependent variable is a function of two or more independent variables. Consider the case of two independent variables, and let an economic illustration take the following form. Gross revenue from sales G can be represented as a function of quantity sold q and price p. We assume for present purposes, what may well not be true in economic reality, that p and q are independent; p is not a function of q, q is not a function of p, and both are not functions of a third, and fundamental, independent variable z.

$$G = pq \tag{50}$$

We may get the rate of change of G with p, the derivative of G with respect to p, as

$$\frac{\partial G}{\partial p} = q$$

and, in doing this, q is treated as a constant because it is assumed that q is *not* a function of, does not "vary with" or "depend upon," p. To emphasize this fact, a new symbol is used for the derivative, involving the ∂ in place of d.

Similarly, the rate of change of G with q is

$$\frac{\partial G}{\partial q} = p$$

If we were interested in the derivatives of the second order, we should find

$$\frac{\partial^2 G}{\partial p^2} = 0$$

since q is assumed *not* a function of p; and

$$\frac{\partial^2 G}{\partial q^2} = 0$$

since p is assumed *not* a function of q.

But suppose that, after getting the first derivative with respect to p,

$$\frac{\partial G}{\partial p} = q$$

we were then interested in the rate of change of this new function with q, *i.e.*, in the derivative of $\partial G/\partial p$ with respect to q.

Obviously

$$\frac{\partial(\partial G/\partial p)}{\partial q} = 1$$

because, for this purpose, q is a variable. The foregoing clumsy symbol is replaced by a new compound symbol, which indicates successive differentiation with respect to the two independent variables[1]

$$\frac{\partial^2 G}{\partial p\,\partial q}$$

It does not make any difference which differentiation, whether that with respect to p or that with respect to q, takes place first, the result for $\partial^2 G/\partial p\,\partial q$ is the same. Thus, from Equation (50),

$$G = pq \tag{50}$$

differentiating first with respect to p and then with respect to q

$$\frac{\partial G}{\partial p} = q$$
$$\frac{\partial^2 G}{\partial p\,\partial q} = 1$$

or, in the other sequence,

$$\frac{\partial G}{\partial q} = p$$
$$\frac{\partial^2 G}{\partial q\,\partial p} = 1$$

This also implies that

$$\frac{\partial^2 G}{\partial q\,\partial p} = \frac{\partial^2 G}{\partial p\,\partial q}$$

and the order in which ∂p and ∂q are arranged in the lower line of the symbol is normally a matter of indifference.

As a more complicated illustration, consider the function

$$y = 5x^3 + 6x^2z + 3xz^2 - 2z^3 + x^2 - 8xz + 3z^2 - 2x + 5z - 9$$

[1] This symbol, like the other derivative symbols already studied, is a compound symbol; it means "the derivative with respect to q of the derivative of G with respect to p." It is a derivative of the second order, because there have been two stages of differentiation; but it differs from the simple second derivative studied above, in that the two stages of differentiation are not with respect to the same independent variable, but with respect to two different variables (assumed independent of each other).

in which x and z are assumed independent of each other. Differentiating first with respect to x

$$\frac{\partial y}{\partial x} = 15x^2 + 12xz + 3z^2 + 2x - 8z - 2$$

and then with respect to z

$$\frac{\partial^2 y}{\partial x\, \partial z} = 12x + 6z - 8$$

or, proceeding in the other order

$$\frac{\partial y}{\partial z} = 6x^2 + 6xz - 6z^2 - 8x + 6z + 5$$

and

$$\frac{\partial^2 y}{\partial z\, \partial x} = 12x + 6z - 8$$

as before.

Even further generalizations of this process of successive differentiation are manifestly possible. If y is a function of more than two independent variables, successive differentiation is possible with respect to each (or any number less than all) of the variables in turn. Thus if y is a function of x, z, s and t, all independent of each other, we can secure

$$\frac{\partial^4 y}{\partial x\, \partial z\, \partial s\, \partial t}$$

and this ultimate result can be obtained by differentiating with respect to the four variables in any order we please. Moreover, in such differentiation with respect to each of several variables, more than one differentiation with respect to any of them can, if needed for the purpose in hand, be included.

Partial and Total Derivatives. If y is a function of two or more independent variables, as in the foregoing, the derivative of y with respect to one of those variables is called a *partial derivative*. Thus, if

$$y = F(x,z)$$

where the symbol $F(\quad)$ means "function of," $\partial y/\partial x$, treating z as constant, is the partial derivative of y with respect to x.

In the preceding section, we have assumed that the two (or more) independent variables are independent of each other and

of a third independent variable t. If now we assume that x and z are functions of some new variable t, the derivative of y with respect to t can be obtained by the following procedure:

$$\frac{dy}{dt} = \frac{\partial y}{\partial x}\frac{dx}{dt} + \frac{\partial y}{\partial z}\frac{dz}{dt}$$

This is called the *total derivative* of y with respect to t. Clearly the partial derivatives, with respect to x and to z, enter merely as aids in getting the total derivative. Thus, if

$$y = ax^2z + bxz^3$$

and

$$x = 3t - 5, \qquad z = 2t^2 + 3t - 1$$

$$\frac{dy}{dt} = 2axz(3) + bz^3(3) + ax^2(4t + 3) + 3bxz^2(4t + 3)$$

and, as here x and z are understood as functions of t, the above may be replaced by a function of the single variable t

$$\frac{dy}{dt} = 6a(3t - 5)(2t^2 + 3t - 1) + 3b(2t^2 + 3t - 1)^3$$
$$+ a(3t - 5)^2(4t + 3) + 3b(3t - 5)(2t^2 + 3t - 1)^2(4t + 3)$$

which could then be simplified. Or, the derivative might be converted to an expression in x and z alone by replacing t by $(x + 5)/3$; but, in this result, x and z would still be understood as functions of t.

For many practical problems in economics, we regard the given independent variables, such as x and z in the foregoing section, as completely independent. We are then concerned only with *partial* derivatives, that with respect to x and that with respect to z. But in cases for which x and z are indirectly linked by being functions of some other variable t, the *total* derivative, of y with respect to t, may prove of central importance in analysis. Even in these cases, however, the partial derivative, with respect to x or to z, still has formal significance and can be calculated by the rules given above.

The Homogeneous Production Function. Suppose that a firm produces a product z by means of a number of factors of production which, for the sake of simplicity, we shall reduce to two. Let their amounts be x and y. The product may then be considered as a function of the quantities of factors applied

$$z = f(x,y)$$

This function represents all the possible combinations of the factors by which any given z can be produced or, as we may also say, all the technological possibilities within the horizon of the firm's management. It is known as the *production function.* The student will readily perceive that this production function will have to fulfill certain conditions if it is to fit the facts. But instead of going into this matter, we shall consider a particular form of it that has played a great role in the literature of theoretical economics during the last half century.

A function $f(x,y)$ of two (or more) independent variables is said to be homogeneous of the mth degree, if, for any number $t \neq 0$,

$$f(tx,ty) = t^m f(x,y).$$

t^m being, of course, a power of t. Of particular importance for economic theory is homogeneity of the first degree, that is to say

$$f(tx,ty) = tf(x,y) = tz$$

Remembering the economic meanings we have attributed to z, x, and y, the student will have no difficulty in realizing that a production function of this kind simply expresses the absence of those economies or diseconomies that may attend a change in the scale of operations or "constant returns to scale"; or, to put it still differently, such a production function means that if the quantities of *all* factors are doubled, trebled . . . , then the quantity of product will also be doubled, trebled . . . . The partiality of many theorists to this type of production function is amply accounted for by the interesting results that follow from it. We shall consider two examples which the student is bound to meet frequently.

First, if a production function is homogeneous of the first degree, both the average productivity of each factor (that is to say, z/x and z/y, the well-known magnitude which, in the case of the factor labor, is called productivity per man-hour) and the *marginal productivity* of each factor (that is to say, $\partial z/\partial x$ and $\partial z/\partial y$) depend only upon the ratio between the amounts of the factors applied and not upon the amounts themselves.[1] Let us prove this for one of the factors, say, x.

[1] The average productivity is a much-discussed magnitude, particularly for the factor labor. Thus, if x is the amount (in man-hours) of labor used in producing output z, z/x is productivity per man-hour.

The number t can be given any value we choose, provided only that it is not zero. For convenience, we choose t equal to $1/x$. Then we have

$$f\left(1, \frac{y}{x}\right) = \varphi\left(\frac{y}{x}\right) = tz = \frac{z}{x}$$

from which

$$z = x\varphi\left(\frac{y}{x}\right)$$

The function f of two variables has been replaced by the function φ of the one variable y/x. Differentiating partially with respect to x, we have

$$\begin{aligned}\frac{\partial z}{\partial x} &= \varphi\left(\frac{y}{x}\right) + x\frac{\partial\varphi(y/x)}{\partial(y/x)}\frac{\partial(y/x)}{\partial x} = \varphi\left(\frac{y}{x}\right) + x\varphi'\left(\frac{y}{x}\right)\left(-\frac{y}{x^2}\right)\\ &= \varphi\left(\frac{y}{x}\right) - \frac{y}{x}\varphi'\left(\frac{y}{x}\right) = F\left(\frac{y}{x}\right)\end{aligned}$$

This proves our proposition. For $\partial z/\partial x$, the marginal productivity of factor x, is seen to equal the difference of two terms: φ which is a function of y/x alone by definition and φ', the derivative of φ, multiplied by y/x. Thus neither y alone nor x alone occurs in the expression for the marginal productivity. Only their ratio y/x enters. Hence we may set $\partial z/\partial x$ equal to some function F of this ratio, as we have done.

Second, if a production function is homogeneous of the first degree, then the shares in the product that will go to the several factors under conditions of perfect competition will exactly exhaust the product whatever the actual scale of operations. Under these conditions the share of every factor, expressed in terms of the product, is equal to the amount employed multiplied by its marginal productivity. In our case, the shares of factors x and y accordingly are

$$x\frac{\partial z}{\partial x} \quad \text{and} \quad y\frac{\partial z}{\partial y}$$

and we have to prove that

$$x\frac{\partial z}{\partial x} + y\frac{\partial z}{\partial y} \equiv z$$

But this is precisely the case, if and only if the production function in question is homogeneous of the first degree. In fact, we have shown that

$$\frac{\partial z}{\partial x} = \varphi\left(\frac{y}{x}\right) - \frac{y}{x}\varphi'\left(\frac{y}{x}\right)$$

Hence

$$x\frac{\partial z}{\partial x} = x\varphi\left(\frac{y}{x}\right) - y\varphi'\left(\frac{y}{x}\right) = z - y\varphi'\left(\frac{y}{x}\right)$$

On the other hand, we have

$$\frac{\partial z}{\partial y} = \frac{\partial\left[x\varphi\left(\frac{y}{x}\right)\right]}{\partial y} = 0 + \varphi'\left(\frac{y}{x}\right)$$

Hence

$$y\frac{\partial z}{\partial y} = y\varphi'\left(\frac{y}{x}\right)$$

and therefore

$$x\frac{\partial z}{\partial x} + y\frac{\partial z}{\partial y} = z - y\varphi'\left(\frac{y}{x}\right) + y\varphi'\left(\frac{y}{x}\right) \equiv z$$

which identity is a special case of what is known as *Euler's theorem.* Fortunately for the so-called *marginal-productivity theory of distribution,* it is not necessary that this relation hold identically (for all values of the variable). It is sufficient that it hold in an equilibrium point.

CHAPTER V

MAXIMA AND MINIMA

The Minimum of a Function. With the derivative available as a tool of analysis, a problem that presented itself earlier (page 37 of Chap. III) can now be attacked with precision. We can now find the precise value of a specific independent variable x, which renders minimum (or maximum, as the case may be) a specific dependent variable y, which is a function of x.

In Chap. III, in our study of the chosen total cost function

$$c = 0.2q^2 + 1.5q + 7 \tag{16}$$

and the corresponding Chart 20, we found that, for a total cost function of that type, at some point L on the cost curve, average cost would be minimum (page 46). Although we showed that the position of L would be determined by the point of contact of a line through O that was tangent to the curve, we were there unable to *calculate* precisely the value of q (or of c) belonging to L. Later, we examined the average cost function derived from Equation (16), and this average cost was

$$\bar{c} = 0.2q + 1.5 + \frac{7}{q} \tag{17}$$

The curve representing this function (Chart 21) appears to have a minimum point L; but, although we could estimate the value of q belonging to this point as approximately 6 (page 47), we had there no means of determining this value precisely. We noted, of course, that the L of Chart 21 corresponds precisely to the L of Chart 20; both represent minimum average cost for the total cost function stated in Equation (16). We are now in a position to determine the q (and thus the c and $\bar{c}$) of L precisely.

In order, however, to present most conveniently the method to be used, we shall for the moment suspend study of Equation (17) and choose instead a fairly simple functional relation between two variables x and y to which we attach no economic meaning.

The function so chosen is represented by a curve that lends itself readily to the kind of geometrical construction needed to explain the method. Let y be a function of x, as

$$y = 0.2x^2 - 1.5x + 7 \tag{51}$$

and the corresponding curve appears in Chart 39. We observe from the chart that y appears to have a minimum value, at a

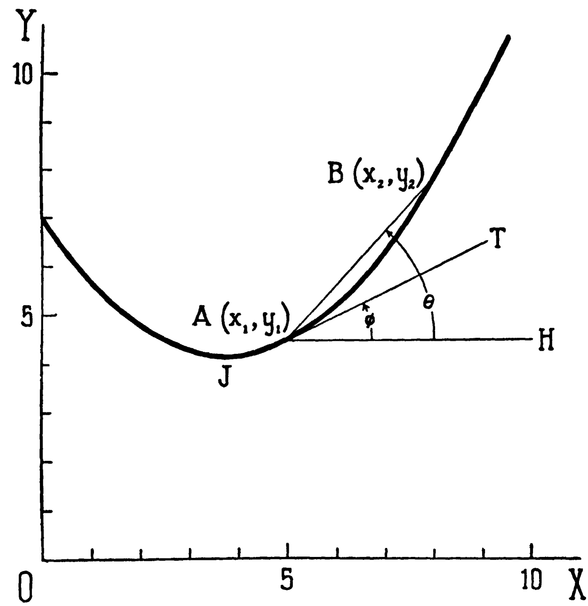

CHART 39.—Graph of $y = 0.2x^2 - 1.5x + 7$, with location of tangent line and minimum point.

point J, for a value of x approximately equal to 4. The immediate task is to determine *precisely* this value of x which renders y a minimum.[1]

Choosing two points $A(x_1,y_1)$ and $B(x_2,y_2)$ of the curve of Chart 39, preferably both to the right of J, and following an analysis similar to that of Equations (16) and (20) to (23) of Chap. III, we find

[1] In this discussion, we shall think of the minimum as the lowest point, and not as a point that is as low as or lower than other points. The more general notion of minimum, which we can exclude here, has considerable theoretical importance.

$$\lim_{\Delta x=0} \frac{\Delta y}{\Delta x} = 0.4x_1 - 1.5$$

as the "instantaneous" rate of change of y with x at A.

But we learned in Chap. IV that

$$\lim_{\Delta x=0} \frac{\Delta y}{\Delta x} = \frac{dy}{dx}$$

and direct application of the rules of differentiation to Equation (51) would have yielded

$$\frac{dy}{dx} = 0.4x - 1.5 \tag{52}$$

for which the value at A, where x is x_i, is

$$0.4x_1 - 1.5$$

Use of the derivative thus obviates the laborious task of imagining a second point B for which x is $x_1 + \Delta x$, making all the computations of Δy and $\Delta y/\Delta x$, and then taking the limit as Δx approaches zero.

However, that lengthy analysis enables us to observe certain important facts. The ratio $\Delta y/\Delta x$ is the slope (see page 12) of the line AB, *i.e.*, of the secant cutting across from point A to point B of the curve (Chart 39). As now, in the limiting process, B slides along the curve toward A, this secant line rotates about point A, taking such successive positions as AB, AB', AB''. As B approaches A, *i.e.*, as Δx approaches zero, the inclination θ of the secant (the angle between a horizontal line AH and the secant AB) becomes smaller. As B gets closer and closer to A, the secant AB approaches a limiting position AT, which is called the *tangent* to the curve at point A. Moreover, the inclination θ approaches a limiting value φ, which is the inclination of the tangent line AT.

We have observed that $\Delta y/\Delta x$ is the slope of AB, and by the relation between slope and inclination, we have also

$$\frac{\Delta y}{\Delta x} = \tan \theta$$

Moreover, as AT is the limiting position of AB, the slope of AT is the limiting value of $\Delta y/\Delta x$ as Δx approaches zero. But this means that the slope of AT is the derivative of y

with respect to x, evaluated for $x = x_1$. Thus

$$\text{Slope of } AT = \frac{dy}{dx}, \quad \text{for} \quad x = x_1$$
$$= 0.4x_1 - 1.5$$

and

$$\tan \varphi = 0.4x_1 - 1.5$$

Furthermore, by taking B left of A but not left of J, a similar analysis could be carried out and would lead to the same result. Here Δx would be negative, and the slope of AT would come out positive as before. Likewise, the analysis can be followed through for A and B left of J; but here the slope of AT, while still equal to the derivative, would come out negative.

We have here a general finding: the slope of the tangent line AT, to a curve at a point A, is the derivative (of the dependent variable with respect to the independent variable) at the point A.[1] This slope of the tangent line at point A is sometimes called the *direction of the curve* at A.

Thus, recapitulating, for the given Equation (51)

$$y = 0.2x^2 - 1.5x + 7$$

we have, by the rules of differentiation,

$$\frac{dy}{dx} = 0.4x - 1.5$$

We can at once calculate the slope of the tangent, the direction of the curve, for any point for which x is chosen. Thus

for x	slope is
5	0.5
8	1.7
6.2	0.98

and, although the graphic analysis took A to the right of J, the results are equally valid to the left of J

for x	slope is
1	-1.1
3	-0.3
2.8	-0.38

[1] It is assumed here that the function is of such form that the derivative "exists" at point A.

We note that, for these latter values of x giving points clearly to the left of J, the slope is negative; hence the tangent of φ is negative, φ is greater than 90°, and the tangent line slopes downward to the right. On the other hand, for the values of x that yield points clearly to the right of J, the slope is positive; hence the tangent of φ is positive, φ is less than 90°, and the tangent line slopes upward to the right.

We may fairly infer that the tangent line is horizontal and its slope is zero, at the lowest point J of the curve. That this conclusion is plausible is indicated by the numerical tests, for selected values of x, made above. A somewhat stronger indication of its truth flows from the following considerations. Suppose A had been taken only very slightly to the right of J and then B had been taken to the right of A. As J is the minimum point (minimum defined in the sense above), the curve must be rising to the right of J; hence, for A and B as chosen, Δy must be positive and $\Delta y/\Delta x$ must be positive. Consequently no matter how close B is to A (so long as it is right of A), Δy is positive and $\Delta y/\Delta x$ is positive; as B approaches A, Δy always remains positive, though it may become exceedingly small, and $\Delta y/\Delta x$ remains positive. Hence the limiting value of $\Delta y/\Delta x$, the slope of AT, is positive.

Now, if A and B are taken to the left of J with B to the left of A, Δx is negative but Δy is positive, because J is a minimum point and the curve is declining (as we pass toward the right along the curve) on the left of J. Each point of the curve is lower than those already passed; hence A is lower than B and Δy is positive. Therefore $\Delta y/\Delta x$ is negative, no matter how close B is to A. Hence in the limit, as B approaches A, the slope of AT is negative. We may imagine this argument followed through even though we took A exceedingly close to J, but still left of J; it would remain true that the slope of AT would be negative.

We infer then that, if J is truly the lowest point, the slope of AT is negative for A left of J and positive for A right of J.[1] We further infer, since at J the slope changes from negative to positive, that the slope of the tangent line at J is zero. This gives

[1] It is assumed here that the curve has only one minimum point and no maximum, as is true of the simple function in Equation (51). If there were other minimum points, this would be called a *relative* minimum (see below, p. 121).

us the key for finding J precisely. As the slope of the tangent at J is zero, the derivative must be zero for that value of x which specifies J. That value of x is therefore found by equating the derivative to zero and solving for x

$$\frac{dy}{dx} = 0.4x - 1.5$$

$$= 0, \qquad \text{for} \qquad x = \frac{1.5}{0.4} = 3.75$$

Thus the exact minimum for y as defined by Equation (51) occurs when x is 3.75, and substitution of this value of x in the original equation gives

$$y = 0.2(3.75)^2 - 1.5(3.75) + 7$$
$$= {}^{335}\!/_{80} = 4.19 \text{ approximately}$$

Thus, point J is (3.75, 4.19).

We now return to Equation (17)

$$\bar{c} = 0.2q + 1.5 + \frac{7}{q}$$

and apply the foregoing method to determine precisely the q belonging to point L of Chart 21. We first differentiate $\bar{c}$

$$\frac{d\bar{c}}{dq} = 0.2 - \frac{7}{q^2}$$

then set this result equal to zero, and solve for q, getting

$$q^2 = 35$$
$$q = 5.916 \text{ (approximately)}^1$$

We estimated, in discussing Chart 21, that q was 6 for L; we now know that the exact value of q at L is $\sqrt{35}$, which is approximately 5.916. We know also that this same q corresponds to point L of Chart 20.[2]

[1] The positive root alone is taken, as the negative root has no economic significance. Actually, Equation (17) is incompletely represented by Chart 21: another *branch* of the curve appears below the negative OQ' axis and left of the negative $O\bar{C}'$ axis, but that branch has no economic meaning.

[2] A less elegant determination of L, resting upon the derivative method but referring to total cost rather than average cost, is also possible. The

The Maximum of a Function. We consider next, for convenience in geometrical analysis, a function somewhat different from that of Equation (51), but one to which we attach no economic significance. Take as this new function

$$y = -0.2x^2 + 1.5x + 7 \qquad (52)$$

Study of the corresponding curve in Chart 40 shows that for some point J y is maximum. As before, we can readily establish that the slope of the tangent line AT, the direction of the curve, at any point A is the derivative of y with respect to x. In this case, we observe that the curve is rising (its direction is positive) to left of J and falling (its direction is negative) to right of J. But the chart suggests, and reasoning like that of the foregoing section indicates, that the tangent line at J is horizontal, its slope is zero. Hence, for a maximum point J, the same condition holds as for a minimum; for the maximum point, the derivative is zero. Therefore to find x for J, we set

$$\frac{dy}{dx} = -0.4x + 1.5$$

---

corresponding total cost function is (p. 42)

$$c = 0.2q^2 + 1.5q + 7 \qquad (16)$$

and study of Chart 20 in Chap. III showed that average cost would be minimum for an output represented by some point L of the curve such that the line OL would be tangent to the curve at L (p. 46). At point L, the slope of OL is a minimum. Let q_1 be the output at L. As OL goes through O, the slope of OL is

$$\frac{c_1}{q_1}$$

where c_1 is given by Equation (16) as

$$c_1 = 0.2q_1^2 + 1.5q_1 + 7$$

Also, the slope of a tangent line at L is the derivative

$$\frac{dc}{dq} = 0.4q + 1.5$$

evaluated for q equal q_1. These two values of the slope must be identical

$$0.4q_1 + 1.5 = \frac{0.2q_1^2 + 1.5q_1 + 7}{q_1}$$

giving

$$0.4q_1^2 + 1.5q_1 = 0.2q_1^2 + 1.5q_1 + 7$$

and from this, once more q_1 is $\sqrt{35}$.

equal to zero and "solve" for x. The result is 3.75 for x, and the corresponding value of y, about 9.81, can be found by substituting this value of x in the original equation which gives y as a function of x.[1]

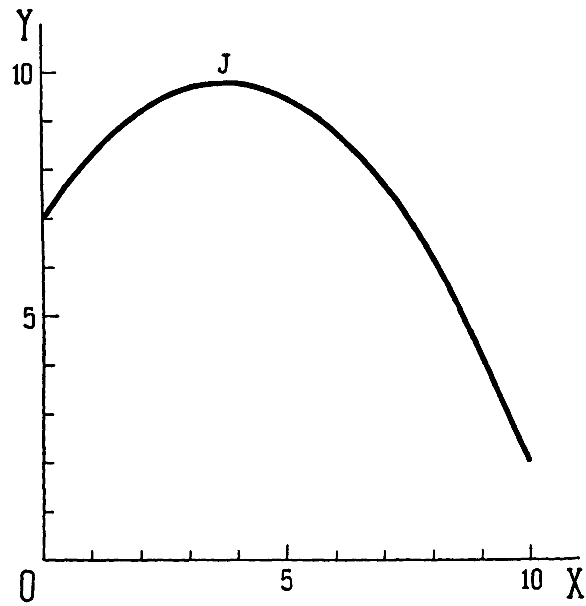

CHART 40.—Curve of a function showing a maximum.

We conclude that either a maximum or a minimum point is yielded when dy/dx is zero. Whether the point is maximum or minimum can readily be inferred from the chart; Chart 39 unmistakably suggests a minimum and Chart 40 equally shows a maximum. In general, this graphic test should be made, especially as it involves no additional labor if the student follows the

[1] That these results give x identical with that of the foregoing section is merely the accidental consequence of the close similarity between the two basic functions. If our basic function in this second instance were

$$y = -3x^2 + 12x - 8$$

we should get

$$\frac{dy}{dx} = -6x + 12, \qquad \text{which} = 0 \text{ for } x = 2$$

and hence J is at (2, 4).

thoroughly wise practice of plotting the curve to secure a visual record of the function.[1]

Distinction between Maxima and Minima. A definite analytical criterion can, however, be set up to tell whether the zero derivative yields a maximum or a minimum. This criterion involves the rate of change in the direction (slope) of the tangent line AT, as A passes from left to right along the curve.

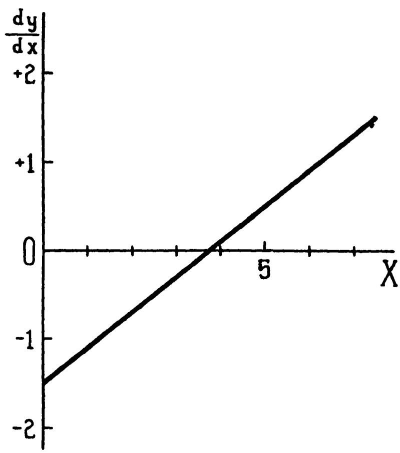

CHART 41.—Derivative of the function of Chart 39.

Referring to Chart 39, the case exhibiting a minimum, a striking feature of the curve is that it is concave upward; its curvature is such that the hollow side lies above the curve. It is like a bowl resting on its base. On the other hand, the curve of Chart 40 is concave downward; its curvature is such that the hollow side is always below the curve. It is like an inverted bowl.

The schedule of slopes (page 113), for various values of x giving various points A of the curve of Equation (51) (Chart 39), shows that the slope is a large negative number for A well to the left of

[1] We developed earlier (p. 69) a utility function

$$u = 80q - \tfrac{4}{15}q^2 \qquad (29)$$

shown in Chart 32 on p. 68. The chart indicates that, for some value of q equal approximately to 150, u is a maximum. From an economic point of view, the significance of u to the right of this point, where u decreases as q increases, is less apparent than to the left, but we may perhaps imagine a situation in which more of a commodity is forced upon the consumer after he has become satiated with it; his utility might then diminish, as q increased beyond the point (maximum u) of satiety.

We proceed now to determine precisely the value of q that makes u of (29) a maximum.

$$\frac{du}{dq} = 80 - \tfrac{8}{15}q$$

and this is zero for $q = 150$. Thus, our above estimate from the chart turns out to be precisely correct.

J, becomes a smaller negative number as A moves right toward J, becomes a small positive number as A passes just to the right of J, and becomes a larger positive number as A passes farther to the right. In other words, having due regard to signs, the slope, *i.e.*, the derivative, appears always to increase as A moves along the curve from left to right.

This is made more emphatically apparent in Chart 41, which represents the derivative of Equation (51). Chart 41 may be regarded as an image of Chart 39, the image relationship being

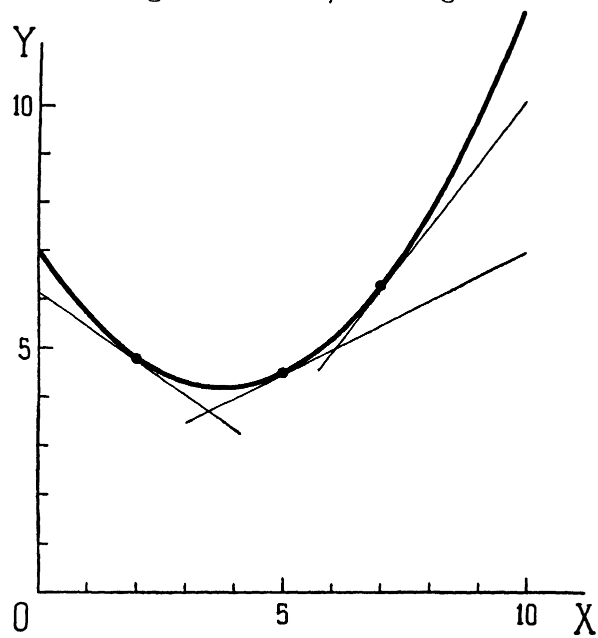

CHART 42.—Tangent lines of various slopes for the curve of Chart 39.

that one chart shows the derivative of the function shown in the other chart. In Chart 41, the "curve" representing the derivative of Equation (51) is a straight line, starts at -1.5 for x equal to zero and rises toward the right, cutting OX for x equal to 3.75. This means that the derivative of the function in Equation (51) always increases as x increases. Moreover, as the curve of Chart 41 is a straight line, the rate of increase of dy/dx is constant.

Remembering that

$$\frac{dy}{dx} = \tan \varphi$$

we can construct, from the formula for the derivative given on page 115, a table of values showing φ for various x (Table 12). As φ is measured counterclockwise from the positive (right) direction of a horizontal line through A, the table shows that the line AT is fairly steep, and sloping downward to the right, for x zero; becomes less steep, but still slopes downward to the right, as x approaches 3.75; is horizontal for x at 3.75; slopes upward to the right for all x greater than 3.75, and becomes increasingly steep as x gets larger after passing 3.75. Several of these tangent lines are ruled on Chart 42, which reproduces the curve of Chart 39.

TABLE 12.—VALUES OF φ, THE INCLINATION OF THE TANGENT LINE TO THE CURVE ON CHART 39, FOR SELECTED VALUES OF x, AS COMPUTED FROM THE DERIVATIVE OF EQUATION (51)

x	$\frac{dy}{dx}$	φ
0	−1.5	123°41′
1	−1.1	132°16′
2	−0.7	145°0′
3	−0.3	163°18′
3.75	0	0°0′
4	0.1	5°43′
5	0.5	26°34′
6	0.9	41°59′
7	1.3	52°26′

The chart and the above considerations indicate that the situation which affords a minimum (in the narrow sense defined above), the situation in which curvature is of the concave-upward type, is one for which the slope of the tangent line (the derivative) dy/dx increases, with due regard to signs, as the point A passes from left to right along the curve.

Without going to the trouble of drawing the image chart (like 41) or calculating various values of φ (like Table 12), we can determine analytically that the slope of the tangent line is increasing. Call the slope m, where

$$m = \frac{dy}{dx} = 0.4x - 1.5$$

To determine whether m is increasing, we need merely to find the rate of change of m with x; if that rate is positive, m increases

with x, if negative, m decreases with x. By the discussion of Chap. IV, page 89, the rate of change of m with x is

$$\frac{dm}{dx}, \text{ which } = 0.4$$

This is positive; hence m increases with x, the curve is concave upward, and J is a minimum and not a maximum.[1]

Furthermore, as m is the first derivative of y with respect to x, dm/dx is the second derivative

$$\frac{dm}{dx} = \frac{d^2y}{dx^2} = 0.4$$

The condition for a minimum may be restated in these terms: The point where dy/dx is zero yields a minimum for y if d^2y/dx^2 is positive (at that point).

By a similar reasoning applied to Equation (52) and Chart 40, we find that the curve, which yields a maximum at J, is concave downward; the slope of the tangent line, with due regard to signs, decreases as A passes from left to right;

$$m = \frac{dy}{dx} = -0.4x + 1.5$$

and

$$\frac{dm}{dx} = \frac{d^2y}{dx^2} = -0.4$$

which is negative. Hence the condition for a maximum may be stated: The point where dy/dx is zero yields a maximum for y if d^2y/dx^2 is negative (at that point). Here, and in the above statement for the minimum, the sign of the second derivative is a sufficient condition for determining the maximum (or minimum); but it is not a necessary condition, as can be shown for certain functions.

Several Extreme Points. A minimum, or a maximum, is an extreme point of the curve; at that point the curve ceases moving vertically, as we pass from left to right, in one direction, downward or upward, and begins to move in the other. The func-

[1] If dm/dx had not been equal to a constant (if y had been a more complicated function of x), but to a variable function of x such as $3x - 8$, we should have needed to find the value of dm/dx for the specific value of x at J and noticed whether the result was positive or negative (see p. 122).

tions considered thus far are of a simple form which yield, in each case, only one extreme point. Functions are readily conceivable, however, which may have a succession of extreme points, alternately maxima and minima.

For example, suppose the function is

$$y = x^3 - 18x^2 + 105x + 10 \tag{53}$$

represented in Chart 43. The extreme points are given when

$$\frac{dy}{dx} = 3x^2 - 36x + 105$$

is zero, *i.e.*, when

$$x = 5 \text{ and } 7$$

As

$$\frac{d^2y}{dx^2} = 6x - 36$$

the first of these points $x = 5$ yields a maximum, and the second $x = 7$ yields a minimum. This last finding is apparent, of course, from the chart as plotted.

With a still more complicated function than that of Equation (53), a larger number of extreme points might appear; for functions of this type, a sum of terms involving positive integral powers of x, the maxima and minima would appear alternately. Cases of this sort can, however, be formulated for which all zero values of dy/dx do not necessarily locate maxima or minima.

In considering a curve having two or more extreme points, such an analysis as that of page 120 and Table 12, if made at all, would need be confined to the vicinity of each single extreme point. In studying variations in φ, *e.g.*, in the vicinity of one of the extreme points (say, the left-hand maximum), we must not allow A (or B) to range up to or beyond the next (adjacent) extreme point.

Points of Inflection. In studying the function of Equation (53), we found the second derivative took the form

$$\frac{d^2y}{dx^2} = 6x - 36 \tag{54}$$

For $x = 5$, this is negative; and for $x = 7$, it is positive. In fact, d^2y/dx^2 is a function of x and therefore changes as x changes. Equation (54) is represented (on a chart with x measured

horizontally and the second derivative measured vertically) by a straight line, inclined upward to the right and cutting OX at

$$x = 6$$

For x less than 6, the second derivative is negative; hence the curve of Equation (53) is concave downward. For x greater than 6, the second derivative is positive; hence the curve of Equation (53) is concave upward. Therefore, for $x = 6$, curvature of the

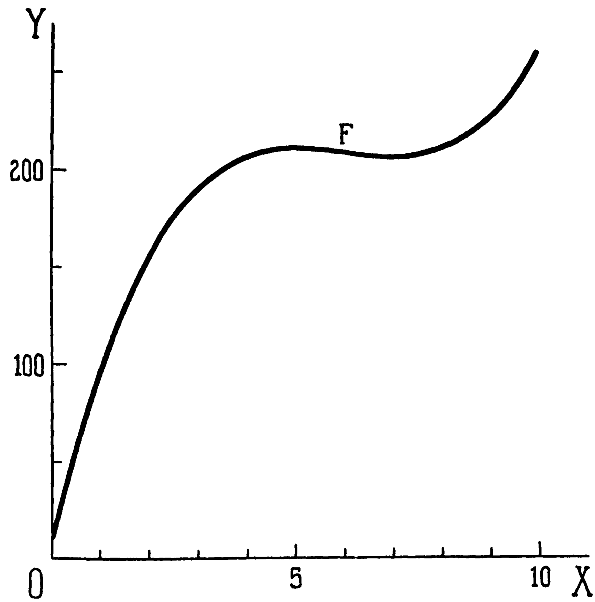

CHART 43.—Function of the cubic-parabola type, with point of inflection.

curve of Equation (53) (Chart 43) shifts from concave downward to concave upward. Such a point, marked F on Chart 43, is called a *point of inflection.*

The provisional condition for determining a point of inflection is thus that

$$\frac{d^2y}{dx^2} = 0$$

In simple cases all one needs do, to find the points of inflection, is to find the second derivative, set it equal to zero, and "solve" for x; but the zero second derivative is not a sufficient condition for a point of inflection, as a properly formulated illustration

would show. For the simplest type of function here shown, there will ordinarily be a point of inflection between each successive pair of maximum and minimum points, and the total number of points of inflection will be one less than the number of maxima and minima combined.[1] Of course, points of inflection may occur in functions that do not display either maxima or minima.

The earlier example studied above, Equation (16), had no points of inflection. For that case the second derivative was a constant and, thus, could not equal zero for any value of q. The sign of the second derivative was therefore always fixed, that is to say, always positive. But in the more complicated functions, such as Equation (53), the sign of the second derivative changes as x changes. Hence the need of the parenthetical phrase "at that point" in stating the criterion, in terms of the second derivative, for distinguishing between maximum and minimum (page 121).

Case of Two Independent Variables. The foregoing analysis for maxima and minima applies to the case in which the given function depends upon a single independent variable. For certain important practical problems, maxima and minima are sought for a function that depends upon more than one independent variable. To illustrate the procedure, only the case of two independent variables will be considered. Suppose the given function is

$$y = F(x,z)$$

where $F(\ )$ means "function of." Both x and z are understood to be independent, and to be independent of each other.

Graphically, such a function would be represented by a surface in three-dimensional space. In the case of a simple function of one variable, the graphic representation yielded a curve in a two-dimensional plane, with the independent variable measured in one direction and the dependent variable measured at right angles to it. Here, the two independent variables x and z are measured at right angles to each other in a plane, and the dependent variable y is measured at right angles to the plane, *i.e.*, along a

[1] The possibility that, even for a fairly simple function, the above conditions will not prove conclusive should be borne in mind. For fuller discussion, consult texts on calculus.

third dimension. Manifestly, if we had three or more independent variables, there would be four or more dimensions altogether, and the graphic figure would involve a four-or-more dimensional space, such a space as is beyond the reach of human experience. Although a solid three-dimensional model can in fact be constructed for the surface in the case of two independent variables, or it can be represented tolerably, by proper use of perspective, upon a two-dimensional plane such as a sheet of paper, graphical aids are imperfect for cases of functions that depend upon three or more independent variables. Even for the three-dimensional, two-independent-variables case, graphic representation frequently takes a simplified form in which various sections of the three-dimensional surface are pictured as curves by assuming selected constant values for one of the variables.

For example, if the given function is

$$y = ax^2 + bz^2 + c$$

where a, b, and c are fixed constants, the three-dimensional surface can be plotted in a solid model if the variables x, z, and y are measured from a single origin O along three mutually perpendicular axes OX, OZ, and OY. We may, however, get in mind the main characteristics of the surface, if we note that fixing the value of one variable, say z, gives us a function of a single variable, gives y as a function of x. Thus,

$$\begin{aligned} &\text{for } z = 0, && y = ax^2 + c \\ &\text{for } z = 1, && y = ax^2 + b + c \\ &\text{for } z = 2, && y = ax^2 + 4b + c \\ &\text{for } z = -1, && y = ax^2 + b + c; \text{ etc.} \end{aligned}$$

Previous study shows (page 32) that each of these simple functions is represented by a curve of the sort illustrated in Chart 44, a simple parabola. Accordingly, all sections of the required surface for which z is constant, sections that are, in other words, parallel to the XOY plane, are parabolas. The various parabolas differ from each other only in that their constant element, c, or $b + c$, or $4b + c$, etc., differs; in this case the constant element is merely the value of y when x is zero.

In like manner, any section of the surface parallel to XOZ can be found by choosing fixed values for y, and any section parallel to YOZ by choosing fixed values for x. This process of describ-

ing a surface piecemeal, by discovering the curves made by selected plane sections across the surface, is laborious and is seldom carried out extensively. Limited use of the process proves, however, an effective aid in disclosing the properties of a

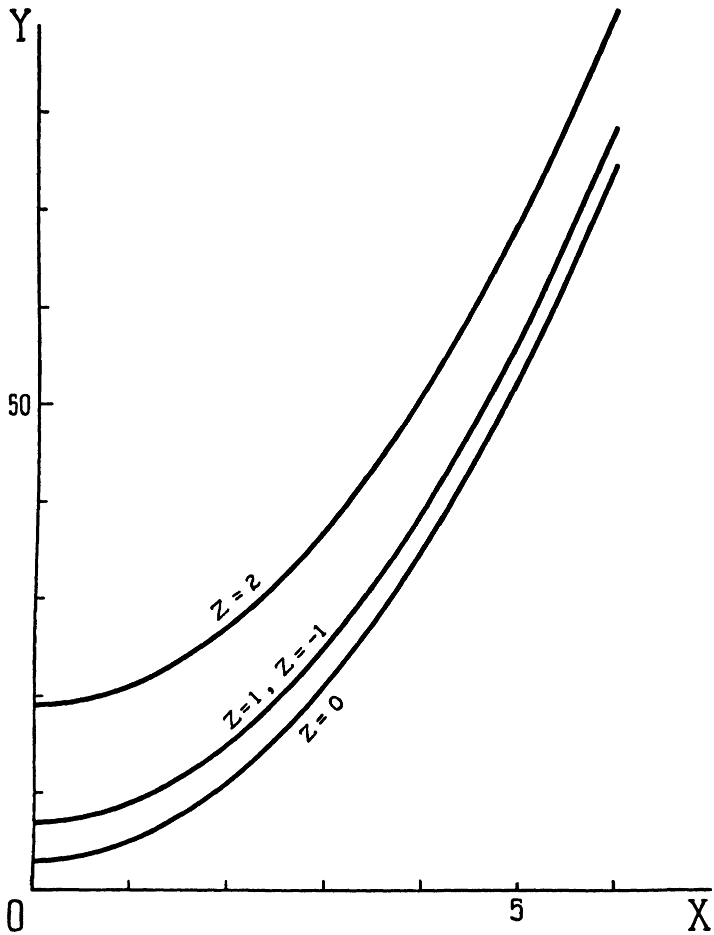

CHART 44.—Sections, parallel to XOY, of a parabolic surface.

three-dimensional figure whether or not we actually make a three-dimensional model or a two-dimensional diagram in perspective of that figure. For the cases of three or more independent variables, involving surfaces in space of more than three dimensions, the only satisfactory graphic tool is that of

representation by sections. The figure in four-dimensional space, for example, is described by taking three-dimensional sections of the surface and thus producing figures that the human eye and mind can visualize.

Minimum Value of a Function of Two Variables. Consider the function of t of two independent variables u and v in the form

$$t = a + gu^2 - hv + kv^2 \tag{55}$$

where a, g, h, and k are fixed constants that will be supposed positive. This function is representable as a three-dimensional surface, the points of which are plotted by measurements of t, u, and v along chosen coordinate axes OT, OU, and OV. A section of the surface parallel to TOU is given for v as some constant, say v_1. Then

$$\begin{aligned} t &= a + gu^2 - hv_1 + kv_1^2 \\ &= a + kv_1^2 - hv_1 + gu^2 \end{aligned} \tag{56}$$

is an equation in the two variables t and u, with $a + kv_1^2 - hv_1$ and g the constants. If these two constants have the numerical values 2 and 1, the curve of the section is as given in Chart 45—it is a parabola. Or, a section parallel to TOV is obtained by choosing a constant u_1 for u, giving

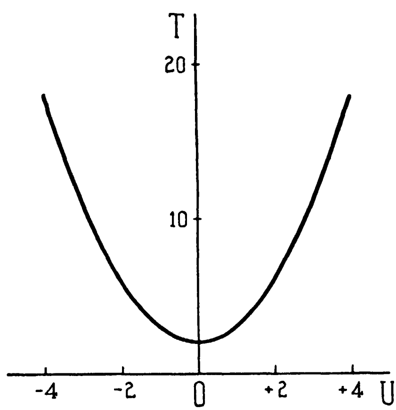

CHART 45.—Section, parallel to UOT, of a parabolic surface.

$$t = a + gu_1^2 - hv + kv^2 \tag{57}$$

which, if charted in the TOV plane, would also show a parabola, but differently located with reference to OT as (57) contains a term in v as well as v^2.

We desire now to find whether t has a minimum (or maximum) value, whether some specified pair of values of u and v yield a minimum (or maximum) for t. So far as the sectional curve [Equation (56)] in a plane parallel to TOU is concerned, we already know how to get its "extreme." Thus

$$\begin{aligned} \frac{dt}{du} &= 2gu \\ &= 0, \qquad \text{for} \qquad u = 0 \end{aligned}$$

and, as

$$\frac{d^2t}{du^2} = 2g$$

is positive (g having been assumed positive) for $u = 0$ (and in fact for all values of u), the point where u is zero yields a minimum for (56). Consider now (57), the equation between t and v.

$$\frac{dt}{dv} = -h + 2kv$$

$$= 0, \qquad \text{for} \qquad v = \frac{h}{2k}$$

and, as

$$\frac{d^2t}{dv^2} = 2k$$

is positive (k having been assumed positive) for $v = h/2k$ (and in fact for all values of v), the point where v is $h/2k$ yields a minimum for (57).

But (56) is the form for t dependent on u alone, when v is any constant, *e.g.*, the constant $h/2k$. And (57) is the form for t dependent on v alone, when u is any constant, *e.g.*, the constant 0. Therefore t is a minimum for the entire surface[1] when u is zero and v is $h/2k$.

We now observe that dt/du for (56) is the same as the partial derivative of t with respect to u for (55)

$$\frac{\partial t}{\partial u} = 2gu$$

Likewise dt/dv for (57) is the partial derivative of t with respect to v for (55)

$$\frac{\partial t}{\partial v} = -h + 2kv$$

In practice, therefore, we determine the minimum (or maximum) of such a function as (55) by taking the two partial derivatives $\partial t/\partial u$ and $\partial t/\partial v$, setting them equal to zero, and solving simul-

[1] This conclusion follows satisfactorily for a function of the exceedingly simple form shown in (55). A more involved mathematical inquiry is needed in determining the extreme of a more complicated function, because the foregoing conditions can be met though no maximum or minimum exists. The reader is referred to texts on calculus for treatment of such cases and is warned that the illustration used here was chosen intentionally to yield a minimum at the point indicated by the conditions.

taneously to find u and v. To test whether the point thus found is maximum or minimum, we observe whether both $\partial^2t/\partial u^2$ and $\partial^2t/\partial v^2$ are negative or positive.[1]

The Lagrange Multiplier. Frequently, problems concerning the existence and location of extreme values of a function of several variables are to be solved subject to a condition. This type of problem can be conveniently dealt with by means of a device that is apt to cause difficulties for students in the perusal of otherwise simple pieces of economic reasoning. An example will illustrate this technique.

Suppose that a firm accepts an order for a definite amount of product $\bar{z}$. The bar is to remind us that in the problem to be discussed the amount of product is not, as it usually is, a variable but a given constant. Also, we shall assume that production involves only two factors x and y and that the prices of these factors, denoted by p_x and p_y, are given to the firm and remain constant during the relevant time. It is required to find that combination of the two factors which under these circumstances will minimize the total cost of filling the order. This cost problem differs from the ones that have been previously treated in this book. Among other things, total cost C is no longer a function of the quantity of product to be produced, which is now a constant, but only of the quantities of the two factors. Moreover, since the firm can choose only among those combinations of factors, or "methods of production," that lie within its horizon or production function (see above, page 107), the minimum of costs it strives to attain can be only a *relative minimum*, *i.e.*, a minimum relative to the possibilities offered by the production function. The latter, therefore, is a condition that the solution will have to fulfill and that plays an essential part in the process of arriving at the solution. Such additional conditions are called *side relations*. Thus we have the cost function

$$C = xp_x + yp_y$$

and the production function as side relation, subject to which total cost is to be minimized

$$\bar{z} = f(x,y)$$

[1] The troublesome cases where these two second partial derivatives are of opposite sign, or where one of them is zero, are ignored here. These will be found treated in calculus texts. See also footnote 1, p. 128.

which we write

$$\bar{z} - f(x,y) = 0$$

We now introduce this side relation into the cost function by means of the following trick. Manifestly, we can always add zero to any number or expression, or subtract zero from any number or expression, without altering its value. We can even, before doing so, multiply the zero with some number λ, because zero multiplied by any finite number is still zero. The expression $[\bar{z} - f(x,y)]$ is equal to zero. Hence nothing prevents us from multiplying it by some number λ, which we shall, however, assume to be positive, and then adding the result to the given cost function. Thus,

$$C = xp_x + yp_y + \lambda[\bar{z} - f(x,y)]$$

This is still a function of the two variables x and y. Its minimum value, if any, can be found by the procedure we have learned in the preceding section. We have

$$\frac{\partial C}{\partial x} = p_x - \lambda \frac{\partial f(x,y)}{\partial x}$$
$$\frac{\partial C}{\partial y} = p_y - \lambda \frac{\partial f(x,y)}{\partial y}$$

These two partial derivatives we now set equal to zero, which evidently yields

$$p_x = \lambda \frac{\partial f(x,y)}{\partial x}$$
$$p_y = \lambda \frac{\partial f(x,y)}{\partial y}$$

From these equations and the given production function, it will ordinarily be possible to eliminate the factor λ, which is commonly referred to as *Lagrange's multiplier*, and to determine the values of x and y that will minimize the total cost of filling the order.[1]

It remains, of course, to make sure whether our result spells an extreme value at all, and whether this extreme value is a mini-

[1] The theory of Lagrange's method is actually not so simple as the exposition in the text might lead the student to believe. It is felt, however, that a common-sense presentation will serve the needs of the beginner better than would a more rigorous one.

mum or a maximum. In our particular case, this is not difficult to find out. As has been seen at the end of the preceding section, we shall have a minimum, if $\partial^2 C/\partial x^2$ and $\partial^2 C/\partial y^2$ are both positive. Now

$$\frac{\partial^2 C}{\partial x^2} = -\lambda \frac{\partial^2 f(x,y)}{\partial x^2}$$

and

$$\frac{\partial^2 C}{\partial y^2} = -\lambda \frac{\partial^2 f(x,y)}{\partial y^2}$$

Since λ is positive by assumption, these second-order partial derivatives of C will be positive, if $\partial^2 f(x,y)/\partial x^2$ and $\partial^2 f(x,y)/\partial y^2$ are both negative. In order to see whether they are or not, we must ascertain their economic meaning. We know already that $\partial f(x,y)/\partial x$ and $\partial f(x,y)/\partial y$ are the marginal productivities of the factors x and y (see page 107). Our second-order partial derivatives are the first partial derivatives of these marginal productivities. The former represent the latter's rates of change with respect to the factor quantities. Therefore, if the second-order partial derivatives are to be negative, the marginal productivities of each factor must decrease—if total product is allowed to vary, then it must increase at a decreasing rate—as further increments of the same factor are added. To express the same thing in familiar economic terminology, increasing inputs of either factor must be attended by decreasing physical returns. This is by no means always the case. But it may be averred that the condition will normally be fulfilled in the region that is relevant for the solution of our problem.

In order to make the economic significance of that solution stand out more clearly, let us return to the equations

$$p_x = \lambda \frac{\partial f(x,y)}{\partial x}$$

$$p_y = \lambda \frac{\partial f(x,y)}{\partial y}$$

and divide the first by the second. We get

$$\frac{p_x}{p_y} = \frac{\partial f(x,y)/\partial x}{\partial f(x,y)/\partial y}$$

The student will have no difficulty in translating this equation into words: The firm must combine the factors in such a way that

the ratio of their marginal productivities equals the ratio of their prices. Our solution therefore yields a general rule for the rational behavior of producers. Alternatively, this rule may be expressed as follows:

$$\frac{1}{p_x}\frac{\partial f(x,y)}{\partial x} = \frac{1}{p_y}\frac{\partial f(x,y)}{\partial y} = \frac{1}{\lambda}$$

In words: in order to minimize total cost of producing a fixed amount of product, the firm must so combine the factors (if it be technologically possible to do so) that the last dollar's worth of any factor should produce the same increment of physical product as does the last dollar's worth of any other factor.

Finally, it should be noticed that λ, which we have introduced as a mere calculatory device, has a definite (though not always the same) economic meaning. Let us see what that meaning is in the particularly simple case of perfect competition.

The quantity of product $z = f(x,y)$ will in this argument be treated as variable. Denote its price by p. Since we are confining ourselves to the case of perfect competition, p is a constant (see pages 2 and 8). For total cost C we use the same expression as above. The net gain π which the firm is striving to maximize is thus

$$\pi = pf(x,y) - xp_x - yp_y$$

For the purpose in hand, we need not go beyond the first step of our procedure, which consists in equating to zero the first partial derivative of π with respect to one of the factors, say, x. This yields

$$\frac{\partial \pi}{\partial x} = p\frac{\partial f(x,y)}{\partial x} - p_x = 0$$

Hence, taking account of the expression for p_x, derived above, we immediately see that

$$p_x = \lambda\frac{\partial f(x,y)}{\partial x} = p\frac{\partial f(x,y)}{\partial x}$$

and so

$$p = \lambda$$

If the factor x is labor, we may express this result as follows: Under conditions of perfect competition the wage rate, the "hourly earnings" of wage statistics, will in equilibrium be

equal to the marginal productivity of (each type of) labor times the price of the product. This theorem is the basis of the so-called "marginal-productivity" theory of distribution.[1]

The Line of Regression. A problem in elementary statistics that involves determination of the minimum of a function of two variables concerns the location of the "line of regression of y on x" for a simple correlation distribution involving two variates x and y (a *variate* x being a particular observed value of the variable x). Suppose that, given a list of such pairs of variates, in considerable number (say n pairs), each x is measured from the mean of all the original values of X and each y is measured from the mean of all the original values of Y. Each individual pair of variates can then be plotted on a *scatter diagram*, in which the x and y are measured along perpendicular axes OX and OY, with the origin O at the mean (Chart 46). Each spot on the chart represents a pair of associated variates; if the spots seem to cluster in an elongated mass, as in the chart, and seem to

[1] The Lagrange multiplier is a convenient and elegant, but not a necessary, device. We shall briefly indicate another procedure. We start again from the cost function

$$C = xp_x + yp_y$$

and the production function

$$\bar{z} = f(x, y)$$

Since $\bar{z}$ is a constant, this production function really expresses a relation between x and y alone. Choosing, arbitrarily, the former for independent variable, we can therefore (ordinarily) write

$$y = \psi(x)$$

and

$$\frac{dy}{dx} = \psi'(x)$$

Now we differentiate totally (see p. 106) both the cost function and the production function with respect to our new and sole independent variable x. Both total derivatives must, of course, be zero, in the case of the cost function in order to fulfill the necessary condition for a minimum, in the case of the production function because $\bar{z}$ is a constant. This yields

$$\frac{dC}{dx} = p_x + p_y\frac{dy}{dx} = 0$$

$$\frac{d\bar{z}}{dx} = \frac{\partial f(x,y)}{\partial x} + \frac{\partial f(x,y)}{\partial y}\frac{dy}{dx} = 0$$

and so, as before

$$\frac{p_x}{p_y} = -\frac{dy}{dx} = \frac{\partial f(x,y)/\partial x}{\partial f(x,y)/\partial y}$$

lie roughly along a straight line that is inclined to the axes, the chart gives provisional evidence of linear association, or *correlation*, between the variables x and y.

Measurement of the degree of such correlation may be approached by determining the lines of regression (one line showing the regression of y on x, the other, of x on y). The line of regression of y on x is defined as a straight line located so that the sum of the squares of the vertical deviations of the

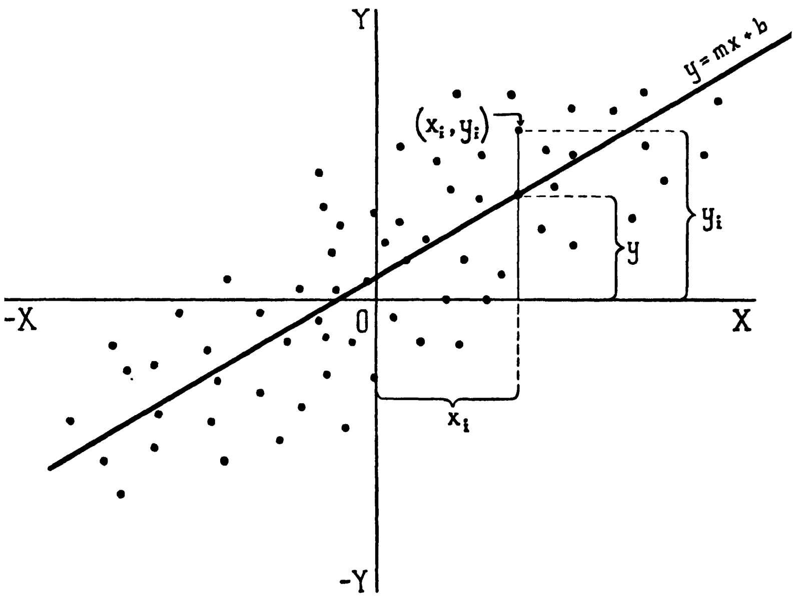

CHART 46.—Correlation scatter and analysis of line of regression.

several spots from that line is a minimum. The general equation of a straight line may be written

$$y = mx + b$$

where m is the slope and b is the y intercept. Consider any spot (x_i,y_i) of the scatter diagram. Its vertical deviation is the difference between y_i and the y of the line at the point vertically below (or above) the spot. For that point of the line, x is manifestly x_i; thus the y of that point of the line is found from the general equation of the line to be

$$y = mx_i + b$$

Hence, the vertical deviation is

$$y_i - (mx_i + b)$$

and its square is

$$[y_i - (mx_i + b)]^2$$

By definition of the line of regression, all these squared vertical deviations, for all spots of the diagram, need to be added, and the total, which we call E, needs to be rendered a minimum. In other words, those values of b and m must be found which will locate the line so that E is a minimum. Now,

$$E = \Sigma\,[y_i - (mx_i + b)]^2$$

where Σ means "the sum of terms like" and this summation extends to include all spots of the diagram, all given pairs of variates. Expanding the expression for E gives

$$\begin{aligned} E &= \Sigma y_i^2 + \Sigma m^2x_i^2 + \Sigma b^2 - \Sigma 2mx_iy_i - \Sigma 2by_i + \Sigma 2mbx_i \\ &= \Sigma y_i^2 + m^2\Sigma x_i^2 + nb^2 - 2m\Sigma x_iy_i - 2b\Sigma y_i + 2mb\Sigma x_i \end{aligned}$$

since factors m^2, $2m$, $2b$, and $2mb$ can be taken "outside" the sign of summation as these factors are constant—do not vary as we pass from spot to spot in the summation process—and since Σb^2 merely means adding together n items equal to b^2. As x_i and y_i are measured from their means,

$$\Sigma x_i = 0 \qquad \text{and} \qquad \Sigma y_i = 0$$

Hence, E reduces to

$$E = nb^2 + \Sigma y_i^2 - 2m\Sigma x_iy_i + m^2\Sigma x_i^2$$

If we introduce three new constants P, σ_x, and σ_y, defined by

$$nP = \Sigma x_iy_i, \qquad n\sigma_x^2 = \Sigma x_i^2, \qquad n\sigma_y^2 = \Sigma y_i^2$$

this last form of E becomes[1]

$$E = nb^2 + n\sigma_y^2 - 2mnP + m^2n\sigma_x^2 \tag{58}$$

Equation (58) gives E as a function of two "variables" b and m; b and m are variable in the sense that, as they change, the position of the line changes. As the problem is concerned with locating the line, this variation of b and m is the means of cover-

[1] The student of elementary statistics will recognize σ_x and σ_y as the standard deviations of x and y, respectively, and P as the product moment.

ing all possible locations, with a view to selecting that location for which E is minimum. All other magnitudes in (58) are constants, so far as the position of the line is concerned; n, σ_y^2, P, and σ_x^2 depend only upon the properties of the original list of variates, and the line can be located in various positions without changing any of these constants.

The next task is to find values of b and m which render E minimum. Following the procedure indicated on pages 127 to 129:

$$\frac{\partial E}{\partial b} = 2nb; \qquad \frac{\partial^2 E}{\partial b^2} = 2n$$

$$\frac{\partial E}{\partial m} = -2nP + 2mn\sigma_x^2; \qquad \frac{\partial^2 E}{\partial m^2} = 2n\sigma_x^2$$

Therefore

$$b = 0 \qquad \text{and} \qquad m = \frac{P}{\sigma_x^2} \tag{59}$$

yield a minimum for E and fix the position of the line of regression[1] of y on x. Once b and m have thus been determined to locate the particular line for which E is minimum, they can be substituted in

$$y = mx + b$$

to yield the equation of the line of regression. In that equation, they (b and m) are constant; the variables there are x and y.

[1] This analysis, in its first stages, is adapted from Yule and Kendall's *An Introduction to the Theory of Statistics*, pp. 209–210, Charles Griffin & Company, Ltd., London, 1937.

CHAPTER VI

DIFFERENTIAL EQUATIONS

Thus far we have been dealing with a supposedly given functional relation, connecting a dependent variable such as c with one or more independent variables such as q. Starting out with such a given functional relation as

$$c = 0.2q^2 + 1.5q + 7$$

we have studied by graphic and other means, and subsequently by the powerful analytical device of the derivative, significant aspects of the relation between c and q.

We now turn to what is in a sense the inverse problem: starting out with given knowledge as to the derivative, in its relation to q or c or both, we seek knowledge concerning the relation between c and q. This inverse process, which involves the concept of the *differential equation* and its analysis, will be found to widen much further our capacity to use analytical machinery for solving economic problems. In many cases the differential equation is not only an exceedingly powerful tool, but also an indispensable tool, for subjecting known or assumed facts to symbolic treatment, and thereby bringing to bear upon them a mathematical process of investigation.

The Differential Equation. An equation that states a functional relation, explicit or implicit, between a derivative (or several derivatives) and the dependent or independent variable or both is called a "differential equation." In contrast, an equation between the variables alone, and not involving any derivatives, may be called an "ordinary equation." The differential equation is thus merely an equation that involves, in one or more of its terms, a derivative (or derivatives).

An example of a simple differential equation giving the derivative explicitly in terms of the independent variable is afforded by any one of the derivative formulas studied above, for instance

$$\frac{dc}{dq} = 0.4q + 1.5$$

This is a simple case in which the derivative is stated equal to an algebraic expression, an expression made up of a sum of terms each of which is a constant multiplied by a positive integral power of the variable, in the independent variable. Its more general form would be

$$\frac{dc}{dq} = a_1q^n + a_2q^{n-1} + \cdot \cdot \cdot + a_nq + a_{n+1} \tag{60}$$

in which n is a positive integer and $a_1, a_2, a_3, \ldots a_{n+1}$ are fixed (or known) constants. As will appear presently, this is a very elementary type of differential equation, which can readily be subjected to symbolic treatment.

An example of a more complicated explicit eauation is

$$\frac{dc}{dq} = ac + bq + k \tag{61}$$

Here both c and q appear in the right expression, and the analysis of such differential equations encounters certain technical obstacles.

As an illustration of an implicit type of differential equation, we may take

$$a\left(\frac{dc}{dq}\right)^2 + b\frac{dc}{dq} = c^2 + q^2 \tag{62}$$

Sometimes, as with an implicit equation of the ordinary sort, an implicit differential equation can be "solved" explicitly for the derivative, thus yielding one of the simpler types mentioned above. In the present instance, Equation (62) can be solved for dc/dq,[1] to yield

$$\frac{dc}{dq} = \frac{-b \pm \sqrt{b^2 + 4a(c^2 + q^2)}}{2a}$$

The right expression here not only differs from the simple algebraic form of Equation (60), but it has the fundamentally more complicated aspect of Equation (61); it involves both the independent variable q and the dependent variable c. Moreover,

[1] Here the solution is written down by use of the formula for the roots of a quadratic equation. The student will recall from his algebra that the formula gives the same result as the more laborious method of "completing the squares."

it is not single-valued; it is really two equations, one for each sign of the radical.

For the time being, we shall ignore all differential equations of the more complicated sort and fix attention upon those of the type represented by Equation (60), an explicit differential equation, in which the derivative is stated equal to an algebraic expression in the independent variable. Fortunately, various important problems in economics lead to differential equations of just such simple type.

Integration. The process of finding the ordinary functional relation, not involving the derivative, between c and q, when the differential equation is given, is called *integration*, or sometimes "finding the solution" or "the integral" of the differential equation. For the exceedingly simple type of differential equation presented in Equation (60), this process is direct and obvious.

Taking, as a specific case of this type, the differential equation

$$\frac{dc}{dq} = 0.4q + 1.5 \tag{63}$$

its solution is found merely by discovering the function c, of q, which when differentiated to yield dc/dq gives exactly the expression on the right side of (63). To discover such a function c, we use the rules of differentiation inversely: we know that the derivative of $0.2q^2$ is $0.4q$, that the derivative of $1.5q$ is 1.5, and that the derivative of *any* constant k is 0. Hence

$$c = 0.2q^2 + 1.5q + k \tag{64}$$

is clearly a function for which dc/dq is $0.4q + 1.5$. Therefore, Equation (64) is the desired solution of Equation (63): Equation (64) gives the ordinary relation between c and q to which the differential equation, given in (63), corresponds.

Two facts should be noted about the process of integration used above. First, we integrated each term in the right member of (63) separately: $0.4q$ yielded $0.2q^2$ as its integral, 1.5 yielded $1.5q$ as its integral. This treatment of the right side of (63) term by term merely amounts to the inverse application of that rule of differentiation (rule 3, page 92) which states that the derivative of the sum of two functions is the sum of the derivatives of the separate functions. The corresponding rule of integration

would read: The integral of the sum of two functions is the sum of the integrals of two separate functions. After taking account of this rule, *each* term of the right side of (63) can then be integrated by an inverse application of the appropriate rule of differentiation. Discovery of which rule is thus appropriate is partly a matter of our familiarity with the rules and our skill in recognizing which rule applies in a particular case; but, as will appear presently, certain technical schemes are of great assistance in those cases in which a search for the appropriate rule might develop into a mere puzzle or process of cut and try.

The second fact to note, and this is of great significance, is that we made allowance in the solution for an additive constant k; this was done on the ground that if there were any additive constant in c its derivative would be zero and would thus be lost from view in (63). This is a very general point: the solution of every differential equation, no matter what other terms (functions of q) it is found to contain, includes an arbitrary constant. This principle applies not only to the simple type of differential equation in (60), but also to the more complicated explicit types.[1] The reason for including the constant in the solution is that, so far as we know, it might be present without altering (since its derivative is zero) the given differential equation.

Such a constant, which we must automatically introduce in the solution of a differential equation, is called a *constant of integration,* or sometimes an *arbitrary constant.* A solution that contains such a constant, as is the case with Equation (64), is called a *general solution* of the differential equation. Manifestly, though all the other constants in the solution may be in numerical form, we do not know the numerical value of the constant of integration. It may have any numerical value (or depend in any specified manner upon other and known constants), without in any way altering the given differential equation; whatever its numerical value, the differential equation remains the same. But different numerical values of the constant—or different specifications of that constant in terms of other (presumably known) constants of the problem, whether numerical

[1] Such a constant enters also even when the differential equation is of the implicit type, but the manner of its appearance may be more intricate than by mere addition.

or otherwise—yield different expressions for the solution, each such expression differing from the others by having a different value of the integration constant. Any one specific value of the constant picks out one specific solution from this whole family of different general solutions; such a specific solution, depending upon a stated value of the constant of integration, is called a *particular solution* of the differential equation.

In practical applications of the method of differential equations, *e.g.*, in the field of economics, we customarily desire a particular solution rather than a general solution. The process of integration, however, yields only the general solution. To pass from this general solution to the desired particular solution, the appropriate value of the constant of integration must be determined. Such determination ordinarily rests upon an *initial condition*—some known pair of associated values of c and q which, when substituted in the general solution, enable us to calculate the constant.[1]

Thus, in the problem to which Equation (63) relates, suppose we knew in advance that c is 10.8 when q is 2; this pair of values is the initial condition. Then, by the general solution (64),

$$10.8 = 0.8 + 3 + k$$

and k is 7. Accordingly, the particular solution is

$$c = 0.2q^2 + 1.5q + 7$$

An equation, like (64), that contains a single arbitrary constant is sometimes said to possess one *degree of freedom.* More complicated differential equations than (63) or (60) can yield through the process of integration general solutions in which several constants of integration appear. If there are m such constants, the general solution is said to possess m degrees of freedom.

Turning now to the more general differential equation of the simple algebraic type

$$\frac{dc}{dq} = a_1q^n + a_2q^{n-1} + a_3q^{n-2} + \cdots + a_nq + a_{n+1} \qquad (60)$$

[1] The term *initial condition* is perhaps unfortunate; for there is nothing necessarily "initial" about it, except in the sense that the condition is known "initially" before and independent of the integration process. But the initial condition need not apply to any beginning point, *e.g.*, the first point of the curve relating c and q on a chart; it can apply to *any* point (see footnote 1, p. 9).

where the various a_i, i ranging from 1 to $n + 1$, are known constants, fixed numerically, and n is a known positive integer, we can at once write down its general solution as

$$c = \frac{a_1}{n+1} q^{n+1} + \frac{a_2}{n} q^n + \frac{a_3}{n-1} q^{n-1} + \cdots + \frac{a_n}{2} q^2 + a_{n+1}q + k \qquad (65)$$

where k is again the constant of integration. As before, the process of integration consists in treating each term of (60) separately, using the appropriate rule of differentiation (mainly the rule for getting dy/dx when y is a power of x) backward, and thus writing down by inspection corresponding terms of (65).

Once more, if a known pair of values of c and q is given as an initial condition, substitution of those values of c and q in (65) determines k and thus yields the desired particular solution.

We have assumed, in Equation (60), that n is a positive integer and, accordingly, that all terms on the right side include only positive integral powers of q. Direct integration, using the same differentiation rules inversely, is nevertheless possible for any powers of q, positive or negative, integral or fractional, except the power minus one. Thus, if

$$\frac{dc}{dq} = aq^{2/3} + bq^{-1/3} + gq^{-4/3}$$
$$c = \tfrac{3}{5}aq^{5/3} + \tfrac{3}{2}bq^{2/3} - 3gq^{-1/3} + k$$

But for the first-negative power of q, a different rule of differentiation must be applied backward—the rule for the derivative of a logarithm. Thus, if

$$\frac{dc}{dq} = \frac{a}{q}$$
$$c = a \log q + k$$

Using a principle of logarithms, this result can be written in the form

$$c = \log q^a + k$$

and, if we replace k by a new constant of integration h related to k by $k = \log h$, this becomes

$$c = \log hq^a$$

Separation of Variables. If both x and y (for which q and c, in the foregoing illustrations, may be regarded as a special case) occur in the expression to which dy/dx is set equal in an explicit differential equation, as in

$$\frac{dc}{dq} = ac + bq + k \tag{61}$$

where k is now a given constant, as are a and b, and not here a constant of integration, the process of integration may prove difficult or impossible. Sometimes, however, the two variables appear in such cases in a form that admits of easy integration by separating the variables. For example, if

$$x^2 \frac{dy}{dx} = (1 - x)y^2 \tag{66}$$

the step preparatory to integration "separates" the variables, by shifting all the factors containing y to the left and those containing x to the right

$$\frac{1}{y^2}\frac{dy}{dx} = \frac{1}{x^2} - \frac{1}{x} \tag{67}$$

This is no longer a differential equation in explicit form, but it can be made so by introducing a new variable z defined so that

$$\frac{dz}{dx} = \frac{1}{y^2}\frac{dy}{dx}$$

This merely implies that z is the following function of y:

$$z = -\frac{1}{y}$$

because application of the rules of differentiation (rule 7, page 93) shows[1] that dz/dx would then in fact be $(1/y^2)\,dy/dx$. Equation (67) can now be written explicitly

$$\frac{dz}{dx} = \frac{1}{x^2} - \frac{1}{x}$$

which integrates to

$$z = -\frac{1}{x} - \log x + k$$

[1] No constant of integration is introduced here, as it would merely amount to having two constants in (68), one on the left and one on the right. These two would then naturally be combined into one—only a single constant really enters.

and, by using the above expression of z in terms of y, this is

$$-\frac{1}{y} = -\frac{1}{x} - \log x + k \tag{68}$$

or

$$\frac{1}{y} = \frac{1}{x} + \log x - k$$

As a matter of fact, k is an unknown constant in the general solution; there is no sense in attaching a minus sign to it, since the sign might be regarded as included in the constant. We can think of a new constant k', equal to $-k$, and then have

$$\frac{1}{y} = \frac{1}{x} + \log x + k' \tag{69}$$

as the general solution. In general, as here and in the $\log hq^2$ case immediately above, we arrange to introduce the constant of integration in whatever form proves most convenient.

In practice, when separation of the variables has thrown the differential equation into such a form as

$$\frac{1}{y^2}\frac{dy}{dx} = \frac{1}{x^2} - \frac{1}{x}$$

we do not customarily take the actual step of introducing the new variable z (see page 155). Instead, we merely note, by the rules of differentiation, applied backward, that the integral of the left side is $-1/y$; then we set down at once

$$-\frac{1}{y} = -\frac{1}{x} - \log x + k$$

As a further example of this procedure, consider the differential equation

$$(1 - y)x\frac{dy}{dx} = y^2 \log x$$

Separating the variables,

$$\frac{1 - y}{y^2}\frac{dy}{dx} = \frac{1}{x}\log x$$

which is

$$\frac{1}{y^2}\frac{dy}{dx} - \frac{1}{y}\frac{dy}{dx} = \log x \frac{d(\log x)}{dx}$$

Integrating,

$$-\frac{1}{y} - \log y = \tfrac{1}{2}(\log x)^2 + k$$

In this case the simplicity of the integration results from the good fortune that, on the right side, $1/x$ is the derivative of $\log x$. In such circumstances—when an expression is itself a derivative of a function, or (as here) consists of two parts one of which is the derivative of the other—the expression is called an *exact derivative.*

Elasticity of Demand. Suppose that the law of demand for a particular commodity in a particular market is specified in the following terms: Elasticity of demand is constant for all amounts of the commodity. This is an illustration of a verbal statement that yields, when thrown into symbolic form, a differential equation, rather than an ordinary functional relation between the fundamental variables (price p and quantity q).

We know that elasticity of demand is given by[1]

$$-\frac{p}{q}\frac{dq}{dp}$$

Hence, for constant elasticity

$$-\frac{p}{q}\frac{dq}{dp} = a \tag{70}$$

where a is a supposedly known constant. Integration of (70) by separating the variables

$$-\frac{1}{q}\frac{dq}{dp} = \frac{a}{p}$$

gives

$$-\log q = a \log p + k$$

or, if we replace k by $-\log h$, involving the new arbitrary constant h, and rearrange the equation

$$a \log p + \log q = \log h$$

[1] If the definition of elasticity has been given, as is sometimes done in elementary economics, in terms of small changes in price and quantity—the elasticity is the ratio of a small relative decrease in quantity corresponding to a small relative increase in price—the derivative form can be obtained by using the limit idea, *i.e.*, letting Δp approach 0 and finding the limit of

$$\frac{\Delta q}{q} \div \frac{\Delta p}{p} \text{ to be } \frac{p}{q}\frac{dq}{dp}$$

giving, by principles of logarithms, the general solution[1]

$$p^a q = h \tag{71}$$

Equation (71) is then the functional relation, representing the demand function, between p and q. It is in implicit form but can be solved explicitly for p or q

$$p = \left(\frac{h}{q}\right)^{1/a} \quad \text{or} \quad q = \frac{h}{p^a}.$$

Here is an example of the standard procedure by which the differential-equation method can be applied to various economic, and other scientific, problems. This situation and procedure are, in outline: (1) the known verbal statement of the problem does not lead to an ordinary functional relation, in symbolic terms, between the variables; (2) it leads rather to a differential equation, which can be written down by a direct translation of the verbal statement into an equation involving one or both variables and the derivative; (3) this differential equation is then integrated, if this is possible from our knowledge of the rules of integration; (4) the arbitrary constant is evaluated from the initial condition, if any is specified in the verbal statement of the problem.

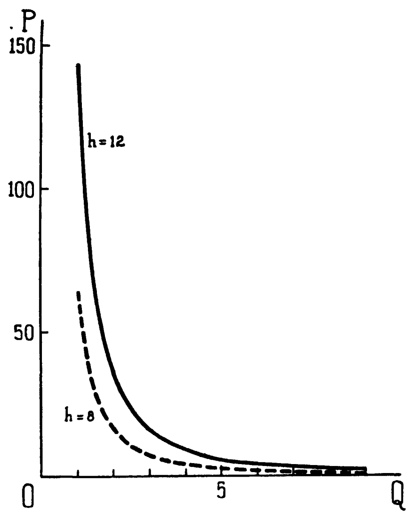

CHART 47.—Selected demand curves of constant elasticity.

For the special case when a is $\frac{1}{2}$, Equation (71) takes the explicit form

$$p = \left(\frac{h}{q}\right)^2$$

and the demand curves, for selected values of h, appear in Chart 47. If a, the elasticity, is unity, Equation (71) becomes

$$pq = h$$

[1] If the verbal statement also informs us as to the initial condition, some known pair of values for c and q, h can be determined specifically, to yield a particular solution.

which means that the total value, pq, of the commodity sold in the market is fixed, whatever the quantity (or price). This case of unit elasticity is especially significant for certain discussions in economic theory. The corresponding curves are equilateral hyperbolas[1] and appear for selected values of h in Chart 48.

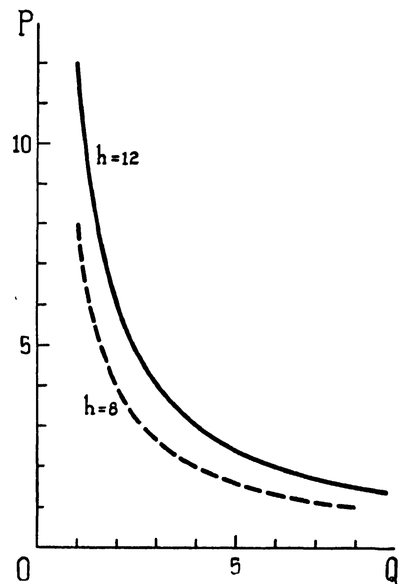

CHART 48.—Selected demand curves of unit elasticity.

Compound Interest. In practical life, compound interest is usually reckoned at stated intervals: a fixed percentage, the "rate of interest," is applied to the accumulated amount of the fund as it existed at the beginning of the interval (say of 6 months) and added thereto at the end of the interval, to yield the new value of the accumulated amount. For certain theoretical purposes, however, the compounding may helpfully be regarded as taking place continuously—the interval may be regarded as an increment of time that approaches zero as a limit.

On this basis, the instantaneous rate of change, with time as the independent variable, of the accumulated fund is the amount of the fund multiplied by the rate of interest. Let t be time, measured from any chosen zero such as the moment when accumulation began, and let F be the accumulated amount of the fund at time t. Then, if the instantaneous rate of interest is the constant r, the verbal conditions of the problem imply the following differential equation:[2]

[1] Only the branch in the positive, q and p both positive, quadrant has economic significance. A similar curve, in the full graphic representation of the mathematical formula, appears diagonally below and to the left, where both q and p are negative.

[2] For clarifying this point, consider a small interval Δt of time t as the interval of compounding. If r is the rate of interest per year, $r\,\Delta t$ will be the percentage to be applied to the fund F, as it existed at the beginning

$$\frac{dF}{dt} = rF \tag{72}$$

By separation of the variables

$$\frac{1}{F}\frac{dF}{dt} = r$$

$$\log F = rt + k$$

and, if k is replaced by $\log h$,

$$\log F - \log h = rt$$

$$\log \frac{F}{h} = rt$$

Using now the definition of logarithm—the logarithm to the base e (and the base e is implied here, see rule 8, page 94) of any number is the exponent of that power of e which equals the given number—this leads to[1]

$$\frac{F}{h} = e^{rt} \qquad \text{or} \qquad F = he^{rt} \tag{73}$$

The compound-interest function, Equation (73), is of high importance in various theoretical problems, in economics and other fields. Although it is seldom or never used in real-life compounding of interest, it enters in numerous ways in financial and other economic problems in which interest can be treated, for theoretical or practical purposes, as being reckoned instantaneously.

To render the formula useful in a given practical case, the rate of interest r would need be known and the constant of integration h would need be determined. Such determination might, for example, flow from a knowledge of the original principal amount P set at interest. This would imply that, for t equal zero, F is P. Then substitution in (73) would show that h is P. The opposite case arises when the accumulated amount A at some future date, t_1 years ahead, is known and the "present value" of the fund,

---

of the interval, to get the interest ΔF to be added at the end of the interval. Thus $\Delta F = Fr\,\Delta t$, and $\Delta F/\Delta t = Fr$, which in the limit gives (72).

[1] This conversion of an equation involving the logarithm function to one involving the exponential function, the function in which e is raised to a variable power, is a device of frequent usefulness. A logarithmic equation in explicit form can always be converted to an exponential equation, and vice versa.

the amount at present that would accumulate to A in t_1 years, is desired. For this case, substituting the initial condition gives

$$A = he^{rt_1} \qquad \text{and} \qquad h = Ae^{-rt_1}$$

and (73) becomes

$$F = Ae^{-rt_1}e^{rt} \tag{74}$$

Under the assumptions as described, the "present" is time zero, and substitution of t as zero in (74) gives F_0 as the present value of the amount A, t_1 years in the future[1]

$$F_0 = Ae^{-rt_1}$$

A related problem, in which the factor corresponding to the "rate of interest" is negative, appears in one type of analysis of depreciation. Suppose that the book value of a capital instrument after time t of its use has elapsed is F and that the instantaneous time rate of depreciation in that value is that value multiplied by some constant $-r$ (where r is positive). Depreciation being a negative change in value, this instantaneous rate is essentially negative (and, if the constant r were used without minus sign attached, r would be negative). Hence

$$\frac{dF}{dt} = -rF$$

and integration gives the general solution

$$F = he^{-rt}$$

If B is the value at the beginning of use, *i.e.*, when t is zero, the constant of integration h is evaluated by

$$B = he^{-rt} = h$$

hence the particular solution is[2]

$$F = Be^{-rt}$$

[1] This result could, of course, have been obtained indirectly by treating the present value F_0 as "principal" P and A as the accumulated value at time t_1.

[2] This is a treatment of depreciation that has certain theoretical advantages, but is not generally used in practice. The commonest method of calculating depreciation in practice is the so-called straight-line method. It deducts a fixed amount from value each year, rather than a fixed percentage: it corresponds to simple interest rather than compound interest.

An interesting point here is that, if the useful life of the machine is also known, as t equal t_1, and the scrap value J at the end of use is also known, r can be determined from this condition.[1] Thus

$$J = Be^{-rt_1}$$

giving

$$\frac{J}{B} = e^{-rt_1}, \qquad \log J - \log B = -rt_1, \qquad \text{and} \qquad r = \frac{\log B - \log J}{t_1}.$$

Total Utility. The measurement, and even the precise definition, of utility does not admit of entirely satisfactory treatment (see page 67). We may, however, make certain rather narrow assumptions upon which a helpful analysis can be developed. Suppose that an individual already has a certain stock q of a consumable commodity A and is considering acquisition of a small additional amount Δq; suppose further that the commodity is available in a market in exchange for money. For present purposes, all that need be assumed about the utility of the commodity to the individual is that the average increase in his utility, per unit of the additional amount Δq, is strictly proportional to the price he is willing to pay.[2] The corresponding symbolic equation is, where u represents utility, p is price, and q is quantity

$$\frac{\Delta u}{\Delta q} = kp, \qquad k \text{ constant}$$

If Δq approaches zero as a limit, this gives the differential equation

$$\frac{du}{dq} = kp$$

and, if the relation between p and q is known from this demand curve as

$$p = f(q)$$

[1] This is not, however, determination of a "constant of integration," only one such constant h enters in integration.

[2] This assumption of strict proportionality is the essential "narrowing" assumption: it is the assumption that may place the present analysis seriously out of touch with reality. It really implies the assumption that the utility of money, which may be regarded as a generalized utility of "other goods," to the individual is not affected by the amount of money spent in exchange for commodity A.

this becomes

$$\frac{du}{dq} = kf(q)$$

Integration of this differential equation, assuming f is a function that can be integrated, yields

$$u = kg(q) + h$$

where g is a new function of q, a function whose derivative is f, and h is the constant of integration. Unless we can assume some such initial condition as that u is zero when q is zero, h cannot be determined. However, although the entire operation presupposes the individual's demand function f is known and makes the somewhat doubtful assumption that k is truly constant (not dependent upon q), this solution of the differential equation assists our understanding of one of the most elusive economic concepts. The individual's marginal utility curve, although expressible in terms of his demand curve, is not identical with it. Once the marginal utility is expressed in terms of the demand function, integration of the differential equation leads to an expression giving the form of the total utility function.

The Differential Notation. Discussion and analysis of a differential equation, such as

$$\frac{dy}{dx} = f(x) \tag{75}$$

can in some respects be facilitated if the equation is written in the form

$$dy = f(x)\,dx \tag{76}$$

and this also is called a differential equation. This amounts, formally at least, to clearing of fractions; dy/dx has been treated as if it were a fraction. We emphasized above (page 88) that dy/dx is not a fraction, but a compound symbol with a meaning that implicitly denies that it is a fraction. Nevertheless, the great convenience of form (76) for certain purposes warrants its use; but, although we treat dy/dx as if it were a fraction in arriving at (76), the student should not forget that the differential equation strictly expresses a functional relation between the derivative, rather than its numerator and denomina-

tor, and the variables. Any operations that we apply to (76) can have validity only if they would also be valid, by implication at least, for the strict form (75) in which the derivative appears in its true guise.

Such a symbol as dx or dy is called a *differential;* more generally, the expression

$$f(x)\,dx$$

may also be called a differential or, better, a differential expression. The analogy between dx and Δx will suggest itself to the reader, and they are treated as identical. For example, if we had a known relation, based upon or derivable from the verbal statement of a particular problem, between the increment Δy in the dependent variable and the corresponding increment Δx in the independent variable in the form

$$\Delta y = f(x)\,\Delta x$$

we should at once write, if $f(x)$ is the derivative of y with respect to x,

$$dy = f(x)\,dx$$

as the differential equation. Here Δy is a function of x and Δx, and likewise dy is a function of x and dx. Although this passage, from a strictly correct equation relating *finite* increments Δy and Δx, to the differential equation is accomplished formally by a mere substitution of dx for Δx and dy for Δy, the entire limit analysis is taken for granted. What we really have done, by implication, consists in (1) dividing Δy by Δx, (2) taking the limit as Δx approaches zero, and then (3) treating the resulting differential equation as if dy/dx could be split up into numerator and denominator like any ordinary fraction. We are in fact defining dy, the differential of y, as the derivative of y with respect to x multiplied by dx.

If the original exact relation between the increments Δy and Δx had taken the form

$$\Delta y = f_1(x)\,\Delta x + f_2(x)\,\Delta x^2 \qquad (77)$$

we might have written

$$dy = f_1(x)\,dx \qquad (78)$$

as the final differential equation. Again, this implies an application of the method of limits. Had we, using (77), divided Δy by Δx, and then allowed Δx to approach zero as a limit, we should have got directly

$$\frac{dy}{dx} = f_1(x)$$

as the differential equation. This, by the definition of dy, could have been thrown into the form (78).

The Concept of the Integral. At the beginning of this chapter, integration has been introduced as the operation that is inverse to differentiation (pages 137 and 138) or as "the process of finding the ordinary functional relation between the dependent and the independent variable" (page 139), when the derivative is given. Also, it has been shown why integration leads to the insertion of an arbitrary constant in the result. Having in the preceding section acquainted ourselves with the differential notation, we are now in a position to introduce the usual symbolism of the integral calculus.

Let a function $f(x)$ be given. If there exists a function $F(x)$ that has $f(x)$ for derivative, then $F(x)$ is called the *integral* of $f(x)$, and the operation of finding $F(x)$ from $f(x)$ is symbolically expressed by the sign $\int$. Thus

$$\int f(x)\, dx = F(x) + c$$

in which the inclusion of dx as a factor on the left side should be noted. By assumption.

$$f(x) = \frac{dF(x)}{dx}$$

The two formulas mean exactly the same thing. It follows that

$$\frac{d[\int f(x)\, dx]}{dx} = f(x)$$

This fact may be expressed by saying that the integral sign and the differential sign cancel each other, just as do square root and second power, $(\sqrt{x})^2 = x$.

To every formula of the differential calculus, therefore, corresponds a formula of the integral calculus. For example

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + c, \qquad n+1 \neq 0$$

$$\int \frac{dx}{x} = \log x + c$$

$$\int \sin x\,dx = -\cos x + c$$

$$\int \cos x\,dx = \sin x + c$$

If in the integral

$$\int f(x)\,dx = F(x) + c$$

we substitute for x first a constant b and then a constant a and subtract, the arbitrary constant c disappears and we have

$$F(b) - F(a)$$

This difference is called the *definite integral* taken between a and b; by way of distinction $\int f(x)\,dx$ is then called the *indefinite integral*. We write

$$\int_{x=a}^{x=b} f(x)\,dx = \Big[F(x)\Big]_a^b = F(b) - F(a)$$

For example

$$\int_0^{\pi/2} \sin x\,dx = \Big(-\cos x\Big)_0^{\pi/2} = -\cos\frac{\pi}{2} - (-\cos 0) = 0 + 1 = 1$$

Exact Differential Equations. Among the advantages of the differential notation is its facilitation of the integration of differential equations in numerous cases: it enables us to proceed without constant thought as to which is the independent variable. For example, consider the differential equation

$$x^2(1-y)\frac{dy}{dx} = (1+x)y^3$$

which, in differential form, is

$$x^2(1-y)\,dy = (1+x)y^3\,dx$$

By separating the variables, this is

$$\frac{1}{y^3}\,dy - \frac{1}{y^2}\,dy = \frac{1}{x^2}\,dx + \frac{1}{x}\,dx \tag{79}$$

The left side of this differential equation can now be integrated, formally at least, without regard to the fact that x is the independent variable. The integral of the left side is

$$-\frac{1}{2y^2}+\frac{1}{y}$$

because the rules of differentiation show that the derivative of this expression, *with respect to y*, is

$$\frac{1}{y^3}-\frac{1}{y^2}$$

The integral of the right side is

$$-\frac{1}{x}+\log x$$

and the solution of the entire differential equation, including k as a constant of integration, is

$$-\frac{1}{2y^2}+\frac{1}{y}=-\frac{1}{x}+\log x+k$$

Any differential expression

$$f(y)\,dy$$

in which $f(y)$ is the derivative—with respect to y, regardless of whether y is or is not the independent variable of the problem, of some other function of y such as $f_1(y)$—*i.e.*, in which

$$f(y)=\frac{df_1(y)}{dy}$$

is called an *exact differential*. The corresponding integral is $f_1(y)$. Thus, in the illustration (left side) of Equation (79),

$$f(y)=\frac{1}{y^3}-\frac{1}{y^2}$$

and the corresponding integral is

$$f_1(y)=-\frac{1}{2y^2}+\frac{1}{y}$$

Likewise, for the right side of (79)

$$g(x)=\frac{1}{x^2}+\frac{1}{x}$$

and the integral is

$$g_1(x) = -\frac{1}{x} + \log x$$

Whenever a differential equation can be thrown into the form involving only exact differentials, this procedure in integration becomes possible. A more complicated type of exact differential involves differentials of both the independent variable and the dependent variable, both dy and dx. Consider, for example, the differential expression

$$2x^3y^2\,dx + x^4y\,dy$$

which can be written in the form

$$x^2y(2xy\,dx + x^2\,dy) \tag{80}$$

If we regard y as constant, the first term

$$x^2y2xy\,dx$$

has the integral

$$\frac{(x^2y)^2}{2}$$

since $2xy$ is the derivative of x^2y with respect to x with y regarded as constant. Considering x constant, the second term

$$x^2yx^2\,dy$$

has the integral

$$\frac{(x^2y)^2}{2}$$

These two integrals are identical; of course, this is a mere fortunate consequence of the peculiar form chosen for the expression (80). Another, and more informing, way of looking at (80) is

$$x^2y\left(2xy + x^2\frac{dy}{dx}\right)dx$$

and, as $2xy + x^2(dy/dx)$ is precisely the derivative of x^2y, this is manifestly an exact differential which might be written

$$x^2y\,d(x^2y)$$

and would therefore have the integral

$$\frac{(x^2y)^2}{2}$$

As noted above, (80) was purposely selected and arranged so that it would be an exact differential. Had it read

$$2x^2y^3\,dx + x^4\,y\,dy$$

it would not have been an exact differential, and its integral could not therefore be found by the foregoing simple process. In actual practice, however, certain expressions do appear which, by the exercise of some ingenuity developed through familiarity with the rules of differentiation, can be thrown into the form of exact differentials. Where this does prove feasible, the method of exact differentials is a powerful aid in integration.

A useful test answers definitely the question whether a differential involving both dx and dy is exact. If the differential is

$$f_1(x,y)\,dx + f_2(x,y)\,dy$$

where f_1 and f_2 are functions of x or y or both, the differential is exact if[1]

$$\frac{\partial f_1}{\partial y} = \frac{\partial f_2}{\partial x}$$

Thus, for (80)

$$f_1 = 2x^3y^2, \qquad f_2 = x^4y$$
$$\frac{\partial f_1}{\partial y} = 4x^3y, \qquad \frac{\partial f_2}{\partial x} = 4x^3y$$

Consider, for example, the differential equation

$$x^5y\,dx + 3xy^2\,dy = 7dx + 2dy$$

which can be written

$$(x^5y - 7)\,dx + (3xy^2 - 2)\,dy = 0$$

Here

$$f_1 = x^5y - 7, \qquad f_2 = 3xy^2 - 2$$
$$\frac{\partial f_1}{\partial y} = x^5, \qquad \frac{\partial f_2}{\partial x} = 3y^2$$

hence the differential equation is not exact. Moreover, the left side of the given equation, considered alone, is similarly shown not exact.

[1] This assumes that f_1 and f_2 meet certain requisites, concerning the existence of their respective derivatives. This point is elaborated in treatises on differential equations.

Integrating Factor. Sometimes the differentials of a differential equation can be rendered exact by multiplying the entire equation by a factor, involving x or y or both, which is called an *integrating factor.* Given the differential equation

$$2y^2\,dx + xy\,dy = 3dx \tag{81}$$

the left side can be made an exact differential if the entire equation is multiplied by x^3 to yield

$$2x^3y^2\,dx + x^4y\,dy = 3x^3\,dx$$

and the essential point is that the right side also remains an exact differential. The left side is now the same as (80) [Equation (81) was purposely chosen, as an illustration, to lead to this outcome], and integration gives

$$\frac{(x^2y)^2}{2} = \tfrac{3}{4}x^4 + k$$

Had the right side of (81) contained dy instead of dx, the integrating factor would have spoiled the right side as an exact differential, and integration would have been halted. The student will enquire whether any method, other than mere inspection or cut and try, exists for finding an integrating factor. If the differential equation can be integrated, the integrating factor can be found; but it cannot always be found in advance. Although, for a limited range of cases, a definite process exists for discovering the integrating factor as a means to integration, no entirely general rule exists; to a large degree the student must rely on his skill in identifying a factor that, in the light of known rules of differentiation, will render a differential exact.[1]

Higher Order Differential Equations. The differential equations that have been examined and integrated above are of a peculiarly simple type: they are *ordinary* differential equations of the first *order* and first *degree.* An ordinary differential

[1] Further discussion of integrating factors will be found in standard works on differential equations. The student is referred to the same sources for information about special devices which have been invented for integrating numerous particular types of differential equations which arise frequently in applied mathematics. Fortunately, many differential equations arising in elementary economics are of the sort discussed in the text above, for which integrals can readily be found.

equation arises only in a case involving two variables, one independent and the other dependent.

The ordinary differential equations so far discussed have involved only *first* derivatives and are called differential equations of the first *order*. A differential equation can arise, however, which involves derivatives of any order, *e.g.*,

$$\frac{d^3y}{dx^3} - 3x\frac{d^2y}{dx^2} + 5x^2\frac{dy}{dx} = 6x^5$$

This is an equation of the third order; in general, the *order* of a differential equation is the same as that of the highest ordered derivative that is included. The integration of differential equations of order higher than the first may be obstructed by serious or even insuperable obstacles, and only a particularly simple case can be taken up here, *viz.*, the case of *linear higher order differential equations with constant coefficients* or differential equations of the general form

$$a_0\frac{d^ny}{dx^n} + a_1\frac{d^{n-1}y}{dx^{n-1}} + a_2\frac{d^{n-2}y}{dx^{n-2}} + \cdots + a_ny = f(x) \qquad (82)$$

where $a_0, a_1, a_2 \ldots$ are constants and $f(x)$ on the right-hand side stands for some expression in the independent variable only. If $f(x)$ is zero, we speak of a homogeneous differential equation of order n, if $f(x)$ is different from zero, of a nonhomogeneous one. We propose to deal with this type of equation by what is known as the *operational* method. For the sake of simplicity we shall, barring a brief remark, confine ourselves to the homogeneous case

$$a_0\frac{d^ny}{dx^n} + a_1\frac{d^{n-1}y}{dx^{n-1}} + a_2\frac{d^{n-2}y}{dx^{n-2}} + \cdots + a_ny = 0 \qquad (83)$$

An *operator* is simply a symbol for a mathematical operation that is to be performed or, as we might also say, a *direction to perform an operation.* $\sqrt{x}$, dy/dx, $\int f(x)\,dx$ are instances that are already familiar to us; the operational symbols involved are $\sqrt{\ }$, d/dx, and $\int$. Now, some of these operators have a property that is most interesting and, at first sight, most surprising. Though they are, of course, not quantities but indications of what we are to do with the quantities to which they refer,

modity (with negligible exceptions) which an individual consumer will take at a stated time (or during a stated interval of time) depends upon the price p of that commodity; (2) that the quantity decreases as the price rises and increases as the price falls, as is in general the case; and (3) that, for most commodities, price does not need to become indefinitely large in order to reduce the consumer's taking to zero, and likewise that even a zero price will not induce the consumer to take an indefinitely large quantity. These characteristics of demand enable us to specify, at least in approximate terms, the functional relation between price and quantity demanded.[1]

TABLE 8.—INDIVIDUAL DEMAND SCHEDULE FOR COMMODITY A. QUANTITIES OF COMMODITY A WHICH A PARTICULAR INDIVIDUAL WILL TAKE, AT A STATED TIME, AT SPECIFIED PRICES*

Quantity	Price	Quantity	Price
†	65	171	35
100	60	200	30
109	55	240	25
120	50	300	20
133	45	400	15
150	40	600	10

Data from GARVER, F. B., and A. H. HANSEN, *Principles of Economics*, p. 100, Ginn & Company, Boston, 1937.

* Units: Cents per "unit," for price; "unit" of commodity, for quantity.

† Not specified.

Although the foregoing description runs in terms of price as independent variable and quantity as dependent variable, a long-standing custom, based largely upon the convenience of fitting various parts of economic theory together, dictates the inverse treatment. As indicated earlier (footnote 1, page 7), which variable is treated as independent and which as dependent is a matter of indifference so far as the mathematical analysis is

[1] The actual finding of such a functional relation, for a given individual consumer and a specified commodity at a stated time, may not admit of completely satisfactory treatment. Broadly, as in the case of the functional relation between cost and quantity (Chap. 1), the problem may be approached either on the a priori or on the empirical basis; but, in practice, serious obstacles would be encountered in either approach. It is sufficient for purposes of the present theoretical discussion to know that such a functional relation exists and that it has some form consistent with the three points given above in the text.

in this case n. Let these roots be denoted by $r_1, r_2, \ldots r_n$.[1] By virtue of a basic theorem of algebra, they may be used for factoring the equation, *i.e.*, for replacing the auxiliary Equation (85) by

$$a_0(m - r_1)(m - r_2)(m - r_3) \cdots (m - r_n) = 0 \qquad (86)$$

which can be satisfied in n ways, *viz.*, by successively equating to zero the factors on the left-hand side

$$\begin{aligned} m - r_1 &= 0 \\ m - r_2 &= 0 \\ \cdots&\cdots \\ m - r_n &= 0 \end{aligned}$$

in each of which m appears linearly.

Therefore, if we now drop the m's, which have served their purpose, and replace them by the D_x's, we have

$$a_0(D_x - r_1)(D_x - r_2)(D_x - r_3) \cdots (D_x - r_n)y = 0$$

and we derive an analogous set of equations, *viz.*,

$$\begin{aligned} (D_x - r_1)y &= 0 \\ (D_x - r_2)y &= 0 \\ \cdots&\cdots \\ (D_x - r_n)y &= 0 \end{aligned}$$

These equations are linear differential equations of the first order and easy to solve. The solution of each contains an arbitrary constant and is a solution of our nth order linear differential Equation (83). Their sum, which will obviously contain n arbitrary constants, is called its *general integral.*

Before illustrating by an example, we shall briefly glance at the procedure to be followed if the right-hand side of an ordinary linear nth-order differential equation with constant coefficients is not zero but a function $f(x)$ of the independent variable. We first disregard $f(x)$, or equate it to zero, and derive the general integral of the remaining homogeneous equation as explained above. In this case, the result is commonly referred to as the

[1] Two or more of them may turn out to be equal. However, since all we can hope to do here is to convey an idea of the principle of the operational method, we shall eliminate this complication by postulating that the roots of the auxiliary equation be all different.

complementary function. Then we try to find a function of the independent variable which, when inserted into the given nth-order equation in place of y, will produce $f(x)$. There are several methods for doing this that are explained in textbooks on differential equations but cannot be presented here. As a final step, the particular integral so found is added to the complementary function. Leaving the matter at this point, we now turn to our example.

Suppose that one of those firms that entered the automobile business during what is known as the "bonanza period" (roughly from 1908 to 1916) experienced a rate of increase in its sales that, at any moment, was proportional to the number of cars already sold. Sales being denoted by S, the time rate of sales was dS/dt, and the rate of increase of this rate of sales was d^2S/dt^2. According to our hypothesis

$$\frac{d^2S}{dt^2} = k^2S$$

or

$$\frac{d^2S}{dt^2} - k^2S = 0$$

k^2 being a positive constant.[1]

It is required to find the integral law that will be satisfied by this ordinary linear second-order differential equation. In order to treat the problem by the operational method,[2] we rewrite

[1] Whenever we wish to emphasize that a constant is positive, we express it by a squared number, because the square of any (real) number is necessarily positive.

[2] In the particular case before us, this is not necessary. In fact, if we multiply both sides of the given equation by $2(dS/dt)$, putting for the sake of convenience

$$\frac{dS}{dt} = u, \qquad \frac{d^2S}{dt^2} = \frac{du}{dt}$$

we have

$$2u\frac{du}{dt} = 2k^2Su$$

The left-hand side now appears as the derivative of u^2, for

$$\frac{du^2}{dt} = \frac{du^2}{du}\frac{du}{dt} = 2u\frac{du}{dt}$$

Replacing u by its value dS/dt on the right-hand side, we get

our equation so as to make it read

$$(D_t^2 - k^2)S = 0$$

The auxiliary equation

$$m^2 - k^2 = (m - k)(m + k) = 0$$

has the roots

$$m_1 = k$$
$$m_2 = -k$$

Therefore we may write

$$(D_t - k)(D_t + k)S = 0$$

which yields the first-order equations

$$D_t S \equiv \frac{dS}{dt} = kS$$

and

$$D_t S \equiv \frac{dS}{dt} = -kS$$

From the first of these equations we derive

$$\frac{dS}{S} = k\,dt$$

which integrates into (see page 142)

$$\log S = kt + \bar{c}$$

or, since $e^{\bar{c}}$ is a constant, say, c_1

$$S = e^{kt+\bar{c}} = e^{kt}e^{\bar{c}} = c_1 e^{kt}$$

From the second of the above equations, we similarly derive

$$\frac{dS}{S} = -k\,dt$$

$$\frac{du^2}{dt} = 2k^2 S \frac{dS}{dt}$$

or

$$du^2 = 2k^2 S\, dS$$

a first-order equation of a type with which we are already familiar (see p. 139). Its integration yields

$$u \equiv \frac{dS}{dt} = \pm\sqrt{k^2S^2 + C}$$

C being the constant of integration. A second integration then solves the problem.

which integrates into

$$\log S = -kt + \bar{c}$$

or, denoting $e^{\bar{c}}$ by c_2,

$$S = e^{-kt+\bar{c}} = c_2e^{-kt}$$

So, finally,

$$S = c_1e^{kt} + c_2e^{-kt}$$

This is the general solution or integral of our equation. But the economic meaning of our argument clearly excludes the second term on the right-hand side; so that we are left with

$$S = c_1e^{kt}$$

which means that the sales of our firm were soaring according to an exponential law (see page 148), a state of things that could not be expected to last and should not have been used for prediction except, perhaps, for the immediate future.

Owing to the frequency of their occurrence, these exponential solutions are recommended to the student's attention. One particularly important point about them should be added. In order to display it, we shall inquire what difference it makes to the general character of the solution of our second-order differential equation if we change the sign that connects its two terms. Suppose that we study the behavior of the price of some highly speculative commodity over a number of months and that our statistical observations suggest the hypothesis that the price, after having been displaced from what we conceive to be its equilibrium position, tends to return to it at a time rate that varies proportionally to its distance from the equilibrium position. Since the tendency in question reduces this distance, the constant factor of proportionality now enters negatively instead of positively as before. Denoting the price by p, its equilibrium value by p_0, the distance between p and p_0 by P, we have

$$\frac{d^2p}{dt^2} = -k^2P$$

or

$$\frac{d^2p}{dt^2} + k^2P = 0$$

or

$$(D_t^2 + k^2)P = 0$$

Proceeding as before, we write the auxiliary equation

$$m^2 + k^2 = 0$$

which yields

$$(m + \sqrt{-k^2})(m - \sqrt{-k^2}) = 0$$

or denoting the "imaginary" number $\sqrt{-1}$ by i,

$$(m + ik)(m - ik) = 0$$

Going on as before, we arrive at the solution

$$P = c_1 e^{ikt} + c_2 e^{-ikt}$$

Now, there exists a relation, known as *Euler's relation*, between exponentials with imaginary exponents and the sine and cosine functions which can only be stated (without proof or discussion) here

$$e^{\pm ix} = \cos x \pm i \sin x$$

Applying this to our solution, we have

$$P = c_1(\cos kt + i \sin kt) + c_2(\cos kt - i \sin kt)$$

or, combining the sine and the cosine terms and introducing new arbitrary constants, A and B, for the sake of simplification

$$P = A \sin kt + B \cos kt$$

The interest of this result proceeds from the fact that our price is practically certain to move in a wavelike or oscillatory manner which may thus find its "explanation." Though economic oscillations are not as a rule of this type, our example, which has no virtue except simplicity, may still serve to give the student a first idea of the nature and use of *dynamic models* in economics.

Higher Degree Differential Equations. An ordinary differential equation of the first *order* is said to be of the second, third, . . . *degree*, if the derivative enters in the second, third, . . . power; the degree of the equation is the same as the highest power to which the derivative occurs. It is possible to deal with this complication by elementary methods if the equation can be solved (1) for the derivative (see page 138), (2) for the independent variable, (3) for the dependent variable. We shall confine ourselves to an extremely simple example of case (3) in order to convey some idea of this kind of problem.

Accordingly, consider the equation

$$y = x\frac{dy}{dx} - \left(\frac{dy}{dx}\right)^3$$

This equation already gives y explicitly, so that we need not solve it. The device to be learned consists in differentiating the equation for x. It is usual to put for convenience

$$p \equiv \frac{dy}{dx}$$

We then have (after differentiating the given equation)

$$\frac{dy}{dx} \equiv p = p + x\frac{dp}{dx} - 3p^2\frac{dp}{dx}$$

Hence, canceling the terms p on each side,

$$0 = (x - 3p^2)\frac{dp}{dx}$$

Manifestly, this equation can be satisfied in two ways. The first is to put

$$\frac{dp}{dx} = 0$$

whence it follows, since the derivative of a constant is zero, that

$$p = c_1$$

Remembering that $p = dy/dx$, we have by a further integration

$$y = c_1x + c_2$$

where c_2 is the *constant of integration* (see page 140). A glance at the given equation tells us that c_2 must be equal to $-c_1^3$. Hence we write (dropping subscript 1)

$$y = cx - c^3$$

But, second, we can also put

$$x - 3p^2 = 0$$

so that

$$p = \pm\sqrt{\frac{x}{3}}$$

and hence, choosing the plus sign,

$$dy = \sqrt{\frac{x}{3}}\,dx$$

or, writing this somewhat differently

$$\sqrt{3}\,dy = x^{1/2}\,dx$$

This integrates into

$$\sqrt{3}y = \tfrac{2}{3}x^{3/2} + c_3$$

or, putting $c_3 = 0$,

$$27y^2 = 4x^3$$

which, as the student may easily satisfy himself by substitution into the given equation, is also a solution.

Partial Differential Equations. In economics, still more than in physics, we have as a rule to deal with problems that involve many independent variables or, at least, more than one. The student will rightly suspect, therefore, that the differential equations of economic theory are likely to contain partial derivatives of various orders and degrees,[1] hence to be *partial* rather than *ordinary* differential equations. The methods of dealing with the former constitute, however, a vast subject into which we cannot enter here.[2] But an important point may be conveyed in an elementary manner.

In Chap. IV (page 104) the following symbol has been introduced:

$$\frac{\partial(\partial G/\partial p)}{\partial q} = \frac{\partial^2 G}{\partial p\,\partial g}$$

the so-called second "cross derivative." Dropping the economic connotation there attached to the letters and replacing the dependent variable G by z, and the independent variables p and q

[1] The definition of linearity (first degree) differs somewhat with different classes of differential equations. Ordinary differential equations and partial differential equations of order higher than the first are called linear if the derivatives and the dependent variable enter in the first powe : but partial differential equations of the first order are called linear if the occurring partial derivatives are linear, no matter what the power of the dependent variable is.

[2] Advanced treatises on calculus usually offer a first introduction to it. There are also special treatises on the subject as well as on subdivisions of it.

by x and y, we can set up a second-order partial differential equation in the simplest possible way by putting that expression equal to zero; thus

$$\frac{\partial^2 z}{\partial x\,\partial y} = 0$$

In order to bring out the significance of this equation, we may use an economic example that, though antiquated, will serve our present purpose. Interpret z as the satisfaction an individual derives from the consumption of the commodities x and y. $\partial z/\partial x$ then stands for the marginal utility of the commodity x to the individual in question and $\partial z/\partial y$ for the marginal utility to him of the commodity y. $\partial^2 z/\partial x\,\partial y$ may, first, be looked upon as the partial derivative of the marginal utility of commodity x with respect to the amount of commodity y, that is to say, as a measure of the rate of the change that occurs in the marginal utility of x, as the individual acquires or loses y. If this change is zero, this manifestly means that such acquisition or loss of y does not affect the marginal utility of x to the individual. But if the marginal utility of x is independent of y, it must depend on x alone and we may write

$$\frac{\partial z}{\partial x} = f(x)$$

or, multiplying both sides of this equation by the differential dx,

$$\frac{\partial z}{\partial x}\,dx = f(x)\,dx$$

Second, $\partial^2 z/\partial x\,\partial y$ may similarly be looked upon as the partial derivative or rate of change of $\partial z/\partial y$, the marginal utility of commodity y, with respect to variations in the quantity of commodity x. If this rate of change is zero, according to the given second-order equation, this again means that the marginal utility of y is independent of the amount of x the individual possesses. Hence the marginal utility of y must depend on the quantity of y alone, and we may write

$$\frac{\partial z}{\partial y} = F(y)$$

or, multiplying both sides of this equation by the differential dy,

$$\frac{\partial z}{\partial y}\,dy = F(y)\,dy$$

By addition of the two foregoing results, we get

$$\frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy = f(x)\,dx + F(y)\,dy$$

Now, at the end of Chap. IV (page 105) the concept of total derivative was introduced. In the light of what has been said in the current chapter (see page 151) on the subject of the differential notation, the reader should have no difficulty in grasping the meaning of the following proposition: If a variable z depends upon two independent variables x and y. its total derivative *with respect to one of them, say x, is*

$$\frac{dz}{dx} = \frac{\partial z}{\partial x} + \frac{\partial z}{\partial y}\frac{dy}{dx}$$

or, "multiplying through" with dx,

$$dz = \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy$$

dz is called the *total differential.* The proposition reads in words: The total increment (positive or negative) that a function of two variables experiences, when both of the latter are allowed to vary simultaneously by "infinitesimal" amounts, may be equated to the sum of two terms, each of which represents the effect on the dependent variable of an increment in one of the two independent variables, the other independent variable being held constant. In physics this is called the "principle of superposition of small effects." Let us apply it to

$$\frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy = f(x)\,dx + F(y)\,dy$$

Since in our example the existence of z is postulated, hence not in question, we evidently may replace the left-hand side of this equation by the total differential dz and write

$$dz = f(x)\,dx + F(y)\,dy$$

or, integrating

$$z = \int f(x)\,dx + \int F(y)\,dy$$

Whatever the properties of the functions f and F may be, they are certainly functions of, respectively, x alone and y alone. Hence the indicated integrations, whatever their particular results may be, cannot produce anything but a function of x alone and a function of y alone, and we derive finally

$$z = \phi(x) + \psi(y)$$

This result tells us two things. First, referring to the economic meaning we have attached to the magnitudes z, x, and y, we have learned that if

$$\frac{\partial^2 z}{\partial x\, \partial y} = 0$$

the marginal and total utilities of our individual's two commodities are independent of each other. In this case, and only in this case, there is sense (granting all the other assumptions implicit in utility analysis) in speaking of the total utility of any commodity taken separately and in adding such total utilities together. The vanishing of our "second cross derivative" has in fact been made the criterion of the independence of commodities by an earlier generation of economists (Edgeworth, Pareto).[1]

But, second, our result also teaches us that the process of integrating or solving partial differential equations will introduce one or more *arbitrary functions* into the result. In our case, f and F, hence also ϕ and ψ, aside from being, respectively, functions of x alone and y alone, are entirely arbitrary, *i.e.*, undetermined so far as the mathematics of the problem is concerned. This is the mathematical fact to which the student's attention should be drawn.

Earlier in this chapter (see pages 139 and 140), it has been pointed out that in solving ordinary differential equations we must introduce arbitrary constants into the solution, the so-called

[1] A possible misunderstanding should be guarded against. Whether two or more commodities are independent or not in the economic sense just explained, they are always independent variables in the mathematical sense. Sugar and coffee are not independent commodities in the economic sense, at least not for most of us: it makes a difference to our enjoyment of a cup of coffee whether or not we have sugar to sweeten it. So far as this is the case, our second cross derivative would not vanish, if x were coffee and y, sugar. But x and y are mathematically independent variables nevertheless: we *can* vary the quantity of the sugar we consume without necessarily also varying the quantity of coffee.

constants of integration. For the case of *additive* constants of this kind, this has been explained on the ground that if there were any additive constant in the integral function to be derived this constant would, the derivative of a constant being zero, fail to show in the differential equation. The fact about partial differential equations that we have just discovered is the exact analogue of this. Only, in the case of partial differential equations, we get arbitrary functions instead of arbitrary constants. Both must be determined from additional information about the problem in hand or, more precisely, about the situation from which the relation starts that is expressed by the differential equation. Marginal costs of a going concern do not contain the element of overhead (fixed or supplementary costs). The function expressing marginal costs, therefore, cannot tell us anything about it, hence the arbitrary constant that enters when we integrate the marginal-cost function, and hence also the arbitrary function that enters the solution of a partial differential equation.

In the conditions of our example, this may be further elucidated as follows. Partially differentiating our solution

$$z = \phi(x) + \psi(y)$$

with respect to x, we get the marginal utility of commodity x

$$\frac{\partial z}{\partial x} = \varphi'(x)$$

As the student observes, we have already lost the function $\psi(y)$ which therefore is entirely arbitrary, provided only that it does not contain x, so far as this first-order partial differential equation is concerned. If now we go on and differentiate once more, this time for y, we have

$$\frac{\partial^2 z}{\partial x\, \partial y} = 0$$

the given second-order partial differential equation from which we started. But *any* functions $\phi'(x)$ and $\phi(x)$ will satisfy it. Since the same result would have followed had we partially differentiated our solution for y instead of differentiating it for x, it is clear that the given second-order equation implies nothing about the form of either ϕ or ψ. Only additional information can tell us what they look like.

CHAPTER VII

DETERMINANTS

In the preceding chapters, the student has been introduced to many functional relations that exist between economic magnitudes such as quantities of commodities, prices, interest rates, and the like. These relations have been expressed in the form of equations, each of which contributes to our understanding of some observed economic facts. But though much can be learned by such analysis of isolated economic relations, we cannot stop at it. In order to understand any given state of the economic organism as a whole or even of any selected part of it, we must also take a comprehensive bird's-eye view of all the relations that are known or are supposed to exist at a given time between our magnitudes and investigate, as it were, their joint operation. This may be illustrated by a very simple example: we learn something by discussing the law of demand of a commodity; we learn something else by discussing its law of supply; but the complete mechanism of a market reveals itself only when we consider both of them simultaneously. This is why, in economics as in other sciences, we have usually to deal with systems of equations rather than with individual equations.

In this simple example the equations to be handled are only two in number. Supposing that they are linear, and denoting price by y and quantity supplied or demanded by x, we may write them in the implicit form (see page 8)

$$\left.\begin{aligned} a_1x + b_1y + c_1 &= 0 \\ a_2x + b_2y + c_2 &= 0 \end{aligned}\right\} \qquad (87)$$

where a_1, b_1, c_1, and a_2, b_2, c_2 are appropriate constants. Solving the equations simultaneously will yield that pair of values of x and y which satisfies both relations. That is to say, solution will yield that price at which quantity supplied and quantity demanded are equal; and that price is particularly interesting for this very reason. Geometrically, this so-called *equilibrium price* is identified by the point of intersection, if it exists, of the

two straight lines that represent the equations. Before proceeding, the student should make sure that he has followed so far.[1]

The Concept of a Determinant. If the systems of equations with which we have to deal were all as simple as the one that has just been considered, there would be nothing more to say. But it is readily seen that the mere number of the relations and conditions that make up the theoretical picture of economic life, even if these relations were all linear, suffices to make the solution of most of the systems of equations that we may be able to construct an extremely laborious if not impossible task. Moreover, we are frequently not interested in the solutions themselves but only in certain properties of them—sometimes even only in the question whether a unique solution exists at all. For both reasons it is important for us to acquaint ourselves with the rudiments of a method that facilitates the solution of more extensive systems of equations or at least provides an easy answer to the question of the existence of a solution, *i.e.*, of a unique set of values of the variables that will satisfy all the equations of the system, provided that the latter is linear. This method consists in the use of *determinants*.

In order to see what a determinant is, we shall first solve equations (87), in the usual way, for the unknowns x and y. We find x by multiplying the first equation by b_2, and the second by $-b_1$, and adding. This yields

$$(a_1b_2 - a_2b_1)x + b_2c_1 - b_1c_2 = 0$$

hence

$$x = \frac{b_1c_2 - b_2c_1}{a_1b_2 - a_2b_1} \tag{88a}$$

Similarly, we find y by multiplying the first equation by $-a_2$, and the second by a_1, and adding

$$(a_1b_2 - a_2b_1)y + c_2a_1 - c_1a_2 = 0$$

[1] The reason why we are particularly interested in the price that equates quantity demanded and quantity supplied is that under certain conditions it is this price that "tends" to prevail and, when departed from, to reestablish itself if the cause responsible for the departure is removed. For this to happen, equality of quantity demanded and quantity supplied is however only a necessary but not a sufficient condition. In addition other conditions, the so-called *secondary conditions*, must be fulfilled. If they are, we speak of *stable equilibrium* which is the real object of our interest. It is not possible to enter into this matter here.

hence

$$y = \frac{c_1a_2 - c_2a_1}{a_1b_2 - a_2b_1} \tag{88b}$$

The student will observe that the denominators of the two fractions are identical and can be obtained by inspection from equations (87) as follows: Multiply the coefficient of x in each equation by that of y in the other, and take the difference between the results. He will observe also that the numerators can be derived from the denominators by replacing, in the first case, the a's and b's by the b's and c's, and, in the second case, the a's and b's by the c's and a's. Knowing these facts, we could have written down the two solutions without actually performing the operations by which we derived them. In our simple case no significant saving in labor would have been achieved by this method. But manifestly this is no longer so when we have to deal with many variables and many equations.

Accordingly, we introduce the following notation:

$$\Delta = a_1b_2 - a_2b_1 = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \tag{89a}$$

$$\Delta_1 = b_1c_2 - b_2c_1 = \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} \tag{89b}$$

$$\Delta_2 = c_1a_2 - c_2a_1 = \begin{vmatrix} c_1 & c_2 \\ a_1 & a_2 \end{vmatrix} = \begin{vmatrix} c_1 & a_1 \\ c_2 & a_2 \end{vmatrix} \tag{89c}$$

The square blocks formed from the constants of our equations are called *determinants* and are usually denoted by the letters D or Δ. Because each of the above determinants consists of two rows and two columns, we call them *determinants of the second order*. Systems of 3, 4, . . . n linear equations in 3, 4, . . . n variables yield square blocks of 3, 4, . . . n rows and columns that are called *determinants* of the 3d, 4th, . . . nth order. For the sake of simplicity, we shall confine ourselves as much as possible to second-order determinants. As we see from (89), they are easily evaluated by cross-multiplication, left-hand upper element times right-hand lower element minus right-hand upper element times left-hand lower element.

Formal Rules about Determinants. In themselves, the determinants in (89), like all determinants of any order, are merely a particular way of writing the algebraic expressions on the left-hand side of (89). Their usefulness is due to the fact that they

can be conveniently handled according to certain mechanical rules. For example, the student can easily verify by evaluation that

1. A determinant's value is zero if one of its rows or columns consists of elements of zero value or if the elements of one row or column are equal or proportional to the elements of the other row or column.

2. A determinant's value remains the same, if rows and columns are interchanged [as shown, for example, by the *two* block arrangements in (89*a*)]; its numerical value remains the same, but its sign is reversed if the two rows or the two columns are interchanged.

3. A determinant's value is multiplied or divided by a constant k if the elements of one row or column are multiplied or divided by k.

These and other rules apply to determinants of any order.

The Solution of Linear Equations. Returning to the solutions for x and y (88*a* and *b*) of the system of equations (87) and taking account of (89), we realize without any further explanation that they can be expressed in terms of the three determinants Δ, Δ_1, and Δ_2. In fact

$$x = \frac{b_1c_2 - b_2c_1}{a_1b_2 - a_2b_1} = \frac{\begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}} = \frac{\Delta_1}{\Delta} \tag{90a}$$

$$y = \frac{c_1a_2 - c_2a_1}{a_1b_2 - a_2b_1} = \frac{\begin{vmatrix} c_1 & c_2 \\ a_1 & a_2 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}} = \frac{\Delta_2}{\Delta} \tag{90b}$$

This result, known as Cramer's rule, is also valid for any system of n nonhomogeneous linear equations in n variables.[1] Consider the system of linear equations

[1] By homogeneous linear equations are meant equations that contain only terms of the form $a_{ij}x_i$, *i.e.*, one of the n variables in the first power multiplied by a constant. Nonhomogeneous linear equations contain, besides such terms, also an additional constant differing from zero. This definition of the term "homogeneous" agrees, as a special case, with the one adopted before (see p. 107).

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = c_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = c_n \end{array}\right\} \tag{91}$$

If the determinant

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{92}$$

is different from zero, then the equations are satisfied by a unique set of values of the variables, *viz.*,

$$x_1 = \frac{\Delta_1}{\Delta}, \qquad x_2 = \frac{\Delta_2}{\Delta}, \quad \cdots \quad x_n = \frac{\Delta_n}{\Delta} \tag{93}$$

where $\Delta_1, \Delta_2, \ldots \Delta_n$ are the determinants that result from the determinant Δ, if we replace the 1st, 2d, . . . nth column, respectively, of the latter by the elements $c_1, c_2, \ldots c_n$.

If

$$\Delta = 0$$

then the system (91) has no solution. The reason and the meaning of this can best be conveyed by a two-variable example. The determinant of the system of equations

$$\left.\begin{array}{l} 3x + 2y = 5 \\ 6x + 4y = 6 \end{array}\right\} \tag{94}$$

is

$$\Delta = \begin{vmatrix} 3 & 2 \\ 6 & 4 \end{vmatrix} = 12 - 12 = 0$$

We see in fact that the two equations (94) are inconsistent, and this is precisely what zero value of the determinant means in this case.

Now let us consider a system of n homogeneous linear equations in n variables, *i.e.*, a system in which all the c's are zero,

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = 0 \end{array}\right\} \tag{95}$$

If we attempt to apply Cramer's rule, we find that all the numerators $\Delta_1, \Delta_2, \ldots \Delta_n$ in the expressions (93) become zero, because they all contain a column of zeros (see rule 1, above). In this case, therefore, the variables can have values different from zero only if $\Delta = 0$. These values are not however unique: the number of solutions of the system is then infinite.

As stated above, it is frequently sufficient for the purposes of economic theory to know whether or not a given system of equation possesses a solution at all, especially a unique solution. The student will observe that the propositions that have been presented enable us to answer this question without actually working out the solution. For instance, in the case of a nonhomogeneous system of n linear equations in n variables we shall have proved the existence of a unique solution as soon as we shall have proved that the determinant of the coefficients is not zero, and this is much easier to establish than it is to solve a system of many equations.

Determinants are also useful for many other purposes. For instance, the student has learned (see, pages 118 to 129) how to distinguish between the maximum and the minimum of a given function in the cases of functions of one variable and of two variables.

By using the differential notation (see page 151), the conditions for the existence of an extreme value of a function of n variables

$$y = f(x_1, x_2, \cdots x_n)$$

may be stated as follows:

1. For an extreme value to exist at the point $x_1 = a_1$, $x_2 = a_2$, $\cdots x_n = a_n$ it is necessary that the total differential (see pages 106 and 169)

$$dy = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} \partial x_2 + \cdots + \frac{\partial f}{\partial x_n} dx_n \tag{96}$$

be zero at that point.

2. For that extreme value to be a maximum value of the function, it is further necessary (and sufficient) that

$$d^2y < 0$$

for all displacements from that point (if we stand on the top of a hill, any possible displacement spells descent); for that extreme value to be a minimum value of the function, it is necessary

(and sufficient) that

$$d^2y > 0$$

for all displacements from that point (if we stand at the bottom of a bowl-shaped valley, any possible displacement spells ascent).[1] If there are many independent variables, d^2y becomes a cumbersome algebraic expression.[2] The use of determinants is then imperative in order to make sure of the sign of d^2y. It is beyond the scope of this chapter to teach this technique. Instead, another technique and some additional concepts pertaining to it will be described below.

The Evaluation of Higher Order Determinants. Since the rules for the evaluation of determinants of any order higher than the third are merely extensions of the rule for the third-order case, we shall confine ourselves to the latter. The basic rule is

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \quad (97)$$

[1] If d^2y is positive for some displacements and negative for others, we have neither a maximum nor a minimum but what is known as a *saddle point*. The case $d^2y = 0$ is not provided for by our criterion as stated above.

[2] The reason for this is that every partial derivative is a function of all the independent variables and hence must be differentiated with respect to all of them. In the case of three independent variables x, y, z, we have

$$d^2f = \frac{\partial^2 f}{\partial x^2}\,dx^2 + 2\frac{\partial^2 f}{\partial x\,\partial y}\,dx\,dy + 2\frac{\partial^2 f}{\partial x\,\partial z}\,dx\,dz + 2\frac{\partial^2 f}{\partial y\,\partial z}\,dy\,dz + \frac{\partial^2 f}{\partial y^2}\,dy^2 + \frac{\partial^2 f}{\partial z^2}\,dz^2$$

The expression on the right-hand side (*any* such expression whatever the number of independent variables) is seen to be of the second degree, a so-called *quadratic form*, in the differentials dx, dy, and dz. From its coefficients (the second-order partial derivatives) the determinants are formed to which allusion is made in the text: the sign of the expression depends upon the signs of those determinants. In our case there are three of them: one of the first, one of the second, one of the third order

$$\left|\frac{\partial^2 f}{\partial x^2}\right|, \qquad \begin{vmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x\,\partial y} \\ \dfrac{\partial^2 f}{\partial x\,\partial y} & \dfrac{\partial^2 f}{\partial y^2} \end{vmatrix}, \qquad \begin{vmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x\,\partial y} & \dfrac{\partial^2 f}{\partial x\,\partial z} \\ \dfrac{\partial^2 f}{\partial x\,\partial y} & \dfrac{\partial^2 f}{\partial y^2} & \dfrac{\partial^2 f}{\partial y\,\partial z} \\ \dfrac{\partial^2 f}{\partial x\,\partial z} & \dfrac{\partial^2 f}{\partial y\,\partial z} & \dfrac{\partial^2 f}{\partial z^2} \end{vmatrix}.$$

It is hoped that the use of the phrase "second order" in two different meanings, second-order derivative and second-order determinant, will not confuse the student.

That is to say, we first take the elements of any row (or column), and multiply each of them by the determinant that we get when we delete from the given determinant the row and the column that intersect in the chosen element. The determinants that we derive by this process of deletion are called *minors* when their absolute value is considered[1] and *cofactors* when a sign is attached to them according to the following rule. The cofactor of any element of a determinant carries a plus sign, when the sum of the numbers (numbering is from the top and from the left) of the row and the column of the chosen element is even, and a minus sign when this sum is odd. In (97), a_{11} stands in row I and column I; I + I = II, hence the cofactor carries a *plus* sign; a_{12} stands in row I and column II; I + II = III, hence the minus sign; and so on.

Using the concept of cofactor, we can say that the second step of the procedure consists in adding the products of the elements of any row or column and their cofactors. The result is the value of the determinant. This is known as the *expansion rule.*[2] In the case of determinants of order higher than the third, the cofactors themselves are determinants of order higher than the second. The expansion rule must then be applied to each of them as often as is necessary in order to reach second-order determinants that can be evaluated by the cross multiplication previously described.[3]

[1] The student will notice that a determinant of order n contains n^2 minors.

[2] By evaluation, the student can easily verify, in the case of (97), the following theorem which is frequently useful in the manipulation of determinants: if the elements of a row or column are multiplied by the cofactors of the elements of any other row and column, then the sum of these products is zero.

[3] Until about 1900, economists made very little use of the determinant notation. Its popularity has steadily grown, however, so that the theoretical literature of the last decade is hardly accessible without some knowledge of the principal methods connected with it. The student who wishes to go beyond this chapter is referred to elementary or advanced treatises on algebra (*e.g.*, to Chap. II of M. Bocher's *Introduction to Higher Aigebra*) and to the special treatises on determinants. All an economist needs in order to understand almost any economic book or paper that uses determinants is assembled in Chaps. XVIII and XIX of R. G. D. Allen's *Mathematical Analysis for Economists*, 1938, perusal of which may perhaps best serve the student as a second step on this road.

INDEX